Sensors Marginalization and Multidimensional Classification for Affective States Recognition

Jesús Joel Rivas, Luis Enrique Sucar Succar, Felipe Orihuela-Espina

Technical Report No. CCC-17-003
July 07, 2017

Luis Enrique Erro 1
Sta. Ma. Tonantzintla,
72840, Puebla, México.

Abstract

Recognizing the affective state that a person is experiencing represents a challenge for computer systems. Some models have been combining information from different sources to mimic how humans simultaneously analyze information from different sensory channels to discern how someone feels. Face expression, body posture, hand gesticulations, and voice tone are external stimuli used to try to estimate the individual's affective state; however there are also sensors that can measure internal physiological reactions that provide more accurate information about the person's affective state; but these sensors are not always available. In machine learning at training phase, all sensors can be used and then at testing or using phase, some of the sensors could be marginalized or set aside. This is the idea under the learning approach using privileged information, but in this case, any of the sensors can be considered privileged information. Additionally, the dependency relationships among affective states can be exploit to enhance the recognition. We hypothesized that a multimodal computational model based on probabilistic graphical models, that allows sensors marginalization and that takes advantage of affective states dependency relationships, can achieve better affective states automatic recognition, having the flexibility of leaving aside any of the sensors in the everyday use of the model. In this work, the developing and validation of a multimodal computational model is proposed for automatic recognition of affective states in the field of interaction with computer games. This model must have the capacity of sensors marginalization and must exploit the dependencies among affective states through multidimensional classification. Identifying the player's affective state can subsequently help to adjust the game to specific needs, with the possibility of applications in virtual environments of rehabilitation and education. The problem is tackled by exploring models with a fusion architecture at decision level using Semi-Naïve Bayesian classifier and/or Bayesian networks, and then using the chaining of such models to make the multidimensional classification, through Bayesian chains classifiers approach. Preliminary experiments reveal the feasibility of the multidimensional computational model to improve the recognition and the marginalization proposal to obtain better results than those obtained with only the available sensors. These models and the algorithms that emerge from it, will represent contributions for affective computing and for computational learning, because of the sensors marginalization and because of the multi-classification that takes advantage of the dependencies of the affective states, at two levels: sensors level and/or classes level. The new models could favour intelligent and empathic human computer interactions.

keywords: affective computing, multimodal systems, marginalization, dependency among affective states, multidimensional classification, probabilistic graphical models.

Contents

Chapter 1
Introduction

In communication between human beings not only the verbal message is transmitted, but also the emotional message of the persons at the moment of communication. The brain analyzes and decodes stimuli from different simultaneous sensory channels, from sources such as facial expression, body posture, hand gesture and/or voice tone; in order to estimate the affective states of the interlocutor. Researchers in affective computing work on the automatic recognition of the user's affective states, by analyzing his/her external behaviour and their physiological records, so the computer can respond accordingly, to promote empathic and intelligent environments that improve human - computer interaction.

1.1 Motivation and justification

Among the challenges reported in affective computing (Picard, 2003; Pantic et al., 2007; Gunes et al., 2008; Zeng et al., 2009; Cernea and Kerren, 2015) for the automatic recognition of emotions are:

1.- The uncertainty around the detection of affective states. Recognition of affective states involves uncertainty inherent in subjective human perceptions both in the process of acquiring the data and in the process of interpreting them. On the other hand affective states are subjective in nature and are variable and influenced by human attributes such as personality, individual's maturity and cultural aspects.

2.- Solutions for detecting affective states should be multimodal. Multimodal in this context means having more than one type of information (having more than one mode, more than one sensor) for the recognition of affective states, for example, facial expressions, physiological signals, body posture, tone of voice, etc. The human interaction, by default, is multimodal. Humans analyze signals from different sensory channels to increase confidence in recognition.

3.- Context should be considered. Aspects such as the geographical location, task(s) that the user is involved, person(s) who the user is interacting with, his/her behaviour in certain situations, etc., can give information about the origin of affective states as triggering factors.

4.- The automatic recognition systems should be feasible to be used in everyday life.

Therefore, it is convenient to propose computational models that consider the associated uncertainty, that incorporate information from different sensors, that identify contextual aspects such as personality and gender of the individual and that it could be used in everyday environments maintaining a good classification performance.

Constructing affective computing multimodal systems that could be used in everyday life, can present some difficulties, if the system includes some sensors for measuring physiological signal. This kind of sensors, e.g., an EEG diadem or EEG cap, the ones for EOG, EMG, ECG, GSR, RSP, TEMP and PLET [1], can be obstructive for the free movement of the user, and some of them can be expensive. However, these physiological signal sensors are more reliable in identifying affective states by taking measurements of the activity of the central and autonomic nervous system (Ekman et al., 1983; Chanel et al., 2011).

Researching in computational models that use information only during training phase and then this information is no longer available during testing or using phase, involves the study of existing classifiers and the creation of mechanisms that allow information marginalization, without deeply degrade the performance of the system. In probability theory the concept of marginalization is used when we have a joint probability distribution $P(X_1, X_2, \cdots, X_n)$ and we want to calculate the marginal distribution of a variable or a subset of them, to obtain an expression or value, where the rest of the variables are marginalized, i.e., they are not present any more. For example, let's suppose we have three random variables: X, Y and Z and their joint probability distribution is $P(X, Y, Z)$, then we want to obtain the joint marginal distribution of X and Y, $P(X, Y) = \sum_Z P(X, Y, Z)$ is calculated and the variable Z is not longer present in the result of $P(X, Y)$, i.e., Z has been marginalized. Marginalization approach can be used to reduce the need of some sensors, consisting in setting as absent any sensor(s) and using the others, and the information previously learned from the absent sensor(s), to obtain as much information as possible to attenuate the loss that the system can suffer at the classification stage. One alternative for marginalization is the paradigm of learning using privileged information (LUPI) which was first introduced by Vladimir Vapnik and Akshay Vashist Vapnik and Vashist (2009) in 2009, and they proposed some solutions by means of the support vector machine (SVM) classifier. For this learning paradigm some information is indicated as privileged information which it is only available at training stage. The potential of this paradigm can be considered for its possibility to register several sensors as privileged information during the training phase (their information are not easily available during the testing phase).

Another element that can contribute to the robustness of the automatic recognition of affective states is that computational models can take into account the dependency relationships among the affective states of interest. Considering co-ocurrent and mutually exclusive relationships among the affective states may improve the discriminative mechanisms for detecting the affective states occurrence (Wang et al., 2015c,b).

The idea of systems that involve sensors marginalization has social impact because people of limited resources could access these systems, and use them in more conventional places, e.g. at home. People could used affective video games which adapt to the user's particular needs, according to his/her affective states (one of the main objectives of the games is to provide emotional experiences (Chanel et al., 2011)). This has important implications in serious games of virtual rehabilitation or education systems, with the idea of promoting patient or student adherence, to the rehabilitation exercises or to the learning of themes, respectively.

[1] EOG: Electrooculography; EMG: Electromyography; ECG: Electrocardiography; GSR: Galvanic Skin Response; RSP: Respiration; TEMP: Skin temperature and PLET: Plethysmograph.

1.2 Problem

In interactive systems of affective computing there are two well-differentiated problems: the recognition of the user's affective states and the decision-making to respond accordingly. If several sources of information from the individual (sensory expressions and/or physiological signals) are analyzed during the automatic user's affective state estimation, a more robust recognition system could be obtained. Additionally, computational models should incorporate the dependency relationships of affective states to reach more discriminating levels among the classes. On the other hand, the user may require to use the system in places where some sensors can be not available. Consequently, it is necessary to propose and develop computational models that analyze user's signals with several sensors while (s)he interacts, for example, with computer games. At first, information should be recorded using physiological signal sensors (for EEG, heart rate, etc.), and devices (they can be seen as sensors) that register external behaviour (facial expressions, hand movement, finger pressure, etc.); afterwards, the models should have the flexibility to allow some of the sensors to be marginalized (to include the case when a sensor is not available) and to consider the dependency relations of the affective states (to have greater discriminative capacity). The marginalization system should embed the sensor's information when it is present at training stage and maintain a higher classification level than the obtained by the mere presence of the other available sensors. The sensor marginalization approach would allow the use of these systems in environments outside the laboratory.

The problem to address consist in proposing a computational model developed under the paradigm of learning using privileged information, which allows the marginalization of sensors and takes advantage of the dependency relations of the affective states involved, to contribute to the recognition of affective states of people who interact with computer games.

1.3 Research questions

Under the assumption of people who interact with computer games, which are induced to experience affective states, and during this process a set of physiological signals and external behavioural information are recorded by a group of sensors:

1.- What marginalization strategies allow to attenuate the performance disturbance of classification (accuracy, F-measure and ROC area) of a system when some of the sensors are removed?

2.- How can be applied the paradigm of learning using privileged information, to make marginalization, in classification techniques from the area of probabilistic graphical models?

3.- How can we exploit affective states dependency relationships to improve automatic affective states recognition in multidimensional classification?

1.4 Research hypothesis

A multimodal computational model based on probabilistic graphical models incorporating sensors marginalization strategies produces better affective states recognition performances (accuracy, F-measure and ROC area) than models which totally ignore and do not use any information (at testing phase) of an indicated (marginalized) sensor (called hereafter trivial marginalization).

1.5 Objectives

1.5.1 Aim

To develop and validate a multimodal computational model for the recognition of affective states, that allows sensors marginalization, to improve classification performance (accuracy, F-measure and ROC area) over the system that does not supply information of the marginalized sensor. Additionally the system will consider the dependency relations of affective states in the multidimensional classification.

1.5.2 Specific objectives

1.- Develop a computational solution for fusing the information, under some scheme, of the sensors of physiological signals and external behavioural in the problem of automatic recognition of affective states.

2.- Evaluate computational techniques that allow the approach of learning using privileged information to develop models of marginalization.

3.- Develop and validate a computational model for the problem of sensors marginalization for automatic recognition of affective states.

4.- Develop and validate a computational model for exploiting the affective states dependence relationships in the multidimensional classification.

5.- Exemplify the utility of the proposed model, applying it to the recognition of affective states of people interacting with a computer game.

1.6 Contributions

1.- A multimodal computational model of affective states recognition that takes advantage of the marginalization, to act when some sensor is absent, achieving classification performance that minimizes the loss of missing sensor information.

2.- The development of a multimodal model for the recognition of affective states that incorporates sensors of physiological signals and external behavioural information; including the hand movements and fingers' pressure.

3.- A multidimensional classification strategy that exploits the dependencies of affective states at two levels: classes level and/or sensors level (you will see a detailed explanation ahead in subsection 4.3.3).

1.7 About preliminary experiments

In a previous work (Rivas et al., 2016) a data set was constructed, which consists of the signals of two stroke patients, during ten rehabilitation sessions over four weeks, on the virtual rehabilitation platform *Gesture Therapy*, with registration of two sensors: hand movement sensor (MOV sensor) and finger's pressure sensor (PRE sensor) for the states of tiredness, anxiety, pain and engagement.

The attribute vectors and class labels: presence or absence of each affective state, resulted in 5826 samples for patient $P1$ and 8935 for patient $P2$; which were labelled by psychiatrists. Three preliminary experiments were carried out on this data set:

1.- The creation of a computational model that receives as input the MOV and PRE sensors and performs late (decision level) fusion through the Semi-Naïve Bayesian (SNB) classifier, this model was called fusion using SNB (FSNB). The classifier for each sensor was proposed in previous work (Rivas, 2015) which was called the Multiresolution Semi-Naïve Bayesian classifier (MSNB), and it is explained in detail at chapter 4,

2.- Multidimensional classification through Bayesian chain classifiers with the models of experiment 1, i.e., FSNB models,

3.- an experiment involving a FSNB model to which a sensor was removed: MOV or PRE, one at a time; The feature values of the missing sensor were estimated with the feature values of the other sensor using simple linear regression.

Details are given in chapter 5. In short, experiment 1 evidenced that the new FSNB computational model produced significantly better results in terms of ROC area with respect to the early (feature level) fusion model of MSNB. ROC area results for the multidimensional classification chain of patient $P2$ were significantly better than the FSNB results of $P2$. Finally, the estimation of the MOV sensor obtained significant results over the model that only has the PRE sensor, for the two patients in terms of ROC area. PRE sensor feature estimation did not improve the results of the model that only has the MOV sensor.

1.8 Organization of the document

Following, in chapters 2 and 3, the theoretical basis is presented, with the fundamental concepts for the development of this work; and the previous works that determine the location of the research and the comparison with the results that would be obtained. Chapter 4 contains the methodological framework and the plan. Chapter 5 describes the preliminary experiments designed to evaluate the feasibility of the proposed solution and the results achieved; and finally the conclusions and future work are indicated in the last chapter.

Chapter 2
Theoretical basis

The theoretical basis is organized in two parts. The first contains the definitions related to affective computing, emotions and personality (sections: 2.1, 2.2, 2.3 and 2.4). The second part contains the set of definitions and main machine learning methods that are referred in this research, including the learning using privileged information paradigm, imputation methods, the Semi-Naïve Bayes classifier, the PKID discretization method, the main techniques of fusion, the multidimensional classification and the Bayesian classifiers in chain, (sections: 2.5 and 2.6).

2.1 Affective states and emotions

For this research we need to know the definitions of affective states, emotion, specific affective states to be treated, personality, affective computing and affective dimensions.

Affective states: They are integrated by emotions, affections, feelings and passions (Betta, 1974). Therefore, the term encompasses all these particular affective manifestations; in fact, the respective affective states are modifications, of different types and intensities, in the habitual humor of an individual: "they are reactive manifestations of the humor provoked by the most diverse external and internal, physical and psychic stimuli" (Betta, 1974).

Emotion: It consists in a neurophysiological response (psychological response with a cognitive evaluation of what is occurring) transient to a stimulus (Matsumoto, 2009), characterized by a "more or less sudden change that occurs in the usual mood or the habitual humor" (Betta, 1974). "The word emotion literally means state of excitement or shudder" (Consuegra, 2010). The highlight element of emotion is the "abruptness of the reaction that provokes in the humor, with the addition of a great physical and psychic repercussion. Of very variable intensity, it arrives, sometimes to motivate and to direct the behaviour that the individual will observe consecutively" (Betta, 1974).

2.2 Affective states and personality

2.2.1 Definition of affective states under consideration

The following section presents the definition of tiredness and affective states relevant to this thesis:

Tiredness: It consists in the sensation of lack of energy that is perceived throughout the body and is not confined to a specific region, producing a decrease in the vitality of the individual. The tiredness emanates from the body and the mind, and it is the interrelation of these two that govern the consciousness of tiredness (Ffrench, 1960).

Anxiety: It is a complex emotion fuelled by not pleasant states of fear, apprehension and, in some cases, anger, linked to expressions of impotence and inability to cope with threatening events (Betta, 1974; Consuegra, 2010). Among its manifestations can be identified the predominance of the physical tension produced by the cognitive part (thoughts, ideas), the physiological part (increased heart rate, altered breathing, dizziness, sweating, tremors and vasomotor changes) and the psychological part (a prolonged sense of tension and prompt preparation before the uncomfortable sensation of imminent danger) (Consuegra, 2010; Bhatia, 2009).

Boredom: It is the emotion that comes when it is not possible to find interests or activities that attract, occupy or completely compromise the individual, arriving at a state of dissatisfaction that makes see things as nonsensical and meaningless (Bhatia, 2009; Consuegra, 2010). It may emerge as a result of external limitations, eg, solitary confinement or sensory deprivation; or as a consequence of internal inhibitions, for example monotomic work (Bhatia, 2009).

Engagement: It is a state of interest (Goldberg et al., 2011), a "positive, fulfilling, and work-related state of mind that is characterized by vigor, dedication, and absorption" (Schaufeli et al., 2002), it includes physical, cognitive and emotional connection to an activity (Hart et al., 2010). It is related to processes that involve information gathering, visual scanning, and lapses of sustained attention (Berka et al., 2007). The individual has a strong intrinsic motivation to achieve a goal (Steinberg, 1996).

2.2.2 Personality

There is no consensual definition of the term personality (Pervin, 2015), possibly there are as many definitions as theoretical and researchers of personality (Wrosch and Scheier, 2003). According to Wrosch and Scheier (2003), personality is "a dynamic organization, within the person, of psychophysical systems that create characteristic patterns of behaviour, thoughts and feelings of the individual". Psychology dictionary Consuegra (2010) defines personality as a "global pattern of behavioural, temperamental, emotional, mental, and character traits that give rise to the unique and relatively consistent way a person feels, think and behave".

The Personality Big Five Factor model (Costa and MacCrae, 1992; Goldberg, 1981, 1990) represents a taxonomy of personality traits in five big dimensions. These dimensions were derived from the analysis of words (originally in English) that people use to describe themselves and others. The model of the big five has served as an integrative framework for different personality description systems and has therefore achieved considerable consensus among researchers in this area (Jeswani and Dave, 2012). The five dimensions or factors of this model are: neuroticism versus emotional stability, extraversion, openness to new experiences, agreeableness (affability or adaptability) and conscientiousness (responsibility or approaches to goals).

Neuroticism: Negative emotionality or neuroticism is a factor that represents the tendency to show a poor emotional adjustment, which leads to negative effects such as anxiety, insecurity and hostility (Jeswani and Dave, 2012).

Extraversion: It is a factor characterized by interest in other people, and the tendency to trust people. This trait promotes sociability, human contact and the ability to talk, care, enthusiasm, assertiveness, energy and active participation (Costa and MacCrae, 1992).

Openness to new experiences: It represents the tendency to be intellectually curious and try to explore new ideas and seek new experiences. Individuals with a high degree of openness to new experiences can be described as creative, imaginative, insightful, innovative, self-reliant, reflective, nonconformist and unconventional (Zhao and Seibert, 2006).

Agreeableness: It consists of the readiness to get along with others; this factor evaluates the orientation of interpersonal relationships. A high affability score reveals a kind, courteous, trusting, caring, forgiving, altruistic, helpful, credulous individual who tries to compromise his/her personal interests with others (Jeswani and Dave, 2012).

Conscientiousness: This factor measures how individuals are able to plan ahead, have the motivation, and direct their impulses toward goal achievement (Costa and MacCrae, 1992).

The model of the Big Five was adapted into Spanish by (Renau et al., 2013; Romero et al., 2012), by translating and organizing the words (personality descriptions) of the language. Different questionnaires have been designed to account for the scores of each of the five factors. These inventories present words organized by row and the individual must indicate how much the respective words identify him. The individual obtains a respective score for each of the five factors and is generally identified with the factor (or the first and second factor) in which he scored the highest.

2.3 Affective computing

Affective Computing (AC) is an area of human computer interaction (HCI), in which solutions are investigated and developed to give computers the ability to recognize human affective states, ideally, at the same level as a person can do (Picard, 2003), and respond accordingly with the possibility of also simulating affective states. Affective computing requires multidisciplinary work, where in addition to computer researchers, the support of experts in the field of psychology is needed.

The users' affective states are not directly observable, but they are expressed through various channels (Luneski et al., 2010, 2008):

1.- writing: contextual information,

2.- auditory: the sound when speaking or expressing,

3.- behaviour: gestures and movements of the face, arms and body in general,

4.- physiological: internal physiological changes (EEG signals, heart beat rate, respiration, blood pressure, dermal conductance, body temperature, etc.).

2.4 Affective dimensions

The study of emotions has led to the creation of models to characterize them. Perhaps one of the most popular is the circumflex model of Russell (Russell, 1980; Posner et al., 2005). In this model a system of Cartesian axes is established to represent, on the abscissa, the **valence** (positive or negative) that extends from negative (unpleasant) levels to positive (pleasant) levels. The axis of the ordinates is used to indicate the level of **arousal** from deactivated (at the bottom) to exited (at the top). Emotions are located in points or zones of this system according to the combination of their valence and arousal levels (Russell, 1980; Posner et al., 2005). The coordinate axes, in this case valence and arousal are known as affective dimensions. Additionally, in other variants a third axis, the **dominance** has been incorporated, to represent the level of control that the person has over the affective situation (Mehrabian, 1980). With the three dimensions, the emotions are represented as points in a three-dimensional space (see figure 2.1).

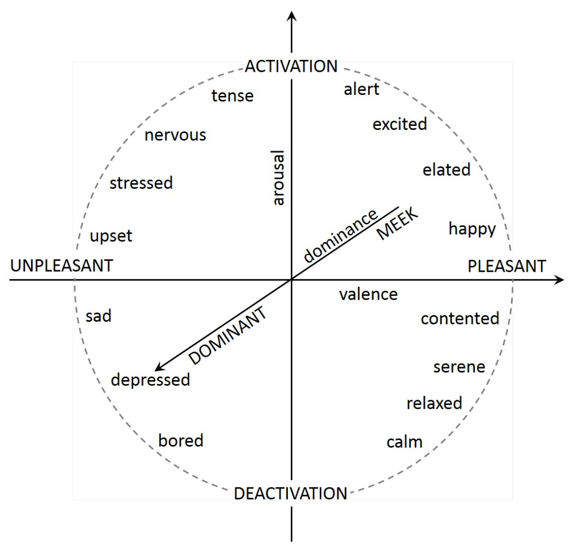


Fig. 2.1: PAD model: valence (pleasure) – arousal - dominance (Mehrabian, 1980). In this model we have three axes: valence, arousal and dominance; emotions are located in points of this three-dimensional space depending on their corresponding levels of the three variables. [modified image from the work Posner et al. (2005)].

2.5 Machine learning methods

2.5.1 Supervised learning

Supervised learning considers the existence of a set of objects Ω. Each object $u \in \Omega$ has two components $u = (\vec{x}_u, c_u)$, where $\vec{x}_u \in \chi$ (usually $\vec{x}_u \in \mathbb{R}^n$) represents the vector of attributes, predictor variables or observable variables, and $c_u \in C$ is the class membership of u (Carrizosa and Martín-Barragán, 2011; Carrizosa et al., 2011). Objects $u = (\vec{x}_u, c_u)$ belong to an unknown distribution $P(\vec{x}_u, c_u)$. There is a non-empty set of objects $E \subset \Omega$, (*training set*). The goal is to predict the class c_v to which an object $v \in \Omega$ is member, knowing only $\vec{x}_v$ (Carrizosa and Martín-Barragán, 2011; Carrizosa et al., 2011); for this, it must be found, from E, within a set of admissible functions $f(\vec{x}_u, \alpha)$ (where parameter $\alpha \in \Lambda$), the function $f(\vec{x}_u, \alpha_*) = c_u\ \forall u$ that minimizes classification error (incorrect values of c_u) (Vapnik and Vashist, 2009).

2.5.2 Learning using privileged information

Learning using privileged information (LUPI) paradigm was originally proposed by Vladimir Vapnik and Akshay Vashist in Vapnik and Vashist (2009). The philosophy of this learning arises when considering the role of a teacher at the moment of teaching a topic. During learning, students receive examples from the teacher (such as supervised learning); but also obtain relevant information such as explanations, comments, comparisons, emphasis on important aspects, and so on. Subsequently, at the test stage, generally, the student has only examples (problems) to apply the knowledge acquired; (s)he no longer has direct access to the additional information provided by the teacher (this privileged information helped him/her to better understand the topic when the teacher explained it).

Formally: Given the objects $u = (\vec{x}_u, \vec{x}^*_u, c_u)$, $u \in \Omega$, obtained from an unknown distribution $P(\vec{x}_u, \vec{x}^*_u, c_u)$, where vector $\vec{x}_u \in \chi$ is known as the observed or available information (features), $c_u \in C$ is the class membership of u and vector $\vec{x}^*_u \in \chi^*$ is the privileged information that is only available during training stage. Given $\alpha \in \Lambda$, the goal consists in finding the parameter α_* for function $f(\vec{x}_u, \alpha_*) = c_u\ \forall u$ to minimize misclassification with help of $\vec{x}^*_u$ (Vapnik and Vashist, 2009).

During training stage, vectors $(\vec{x}_u, \vec{x}^*_u, c_u)$ are supplied; which contain the privileged information, instead of supplying the pairs $(\vec{x}_u, c_u)$ as in the classical paradigm of supervised learning. The additional Information $\vec{x}^*_u \in \chi^*$ belongs, in general, to space χ^* which is different to space χ (Vapnik and Vashist, 2009).

2.5.3 Imputation methods

Privileged information can be seen as missing information at testing stage. When dealing with missing values, imputation is usually employed to compensate the missing data. Imputation is defined as the practice of replacing missing data with substituted values; the values are filled-in (or imputed) using some methods (Little and Rubin, 2014; Schafer and Graham, 2002). Indeed, imputations include outlines of a predictive distribution of the missing values, and require some method for creating a predictive distribution based on the observed data (Little and Rubin, 2014). Most imputation methods are based upon the problem of estimation of the mean and the variance of the data related with the considered value (Muñoz Rosas and Alvarez Verdejo, 2009). There are two generic approaches for generating the predictive distribution (Little and Rubin, 2014):

Explicit modeling: The predictive distribution is based upon a formal statistical model, and hence the assumptions are explicit. These methods correspond to (a) Mean imputation, (b) Regression imputation and (c) Stochastic regression imputation.

Implicit modeling: They are focused on an algorithm, which implies an underlying model;

assumptions are implicit. Some methods are: (a) Hot deck imputation, (b) Substitution, (c) Cold deck imputation and (d) Composite methods.

2.6 Classification techniques

2.6.1 Naïve Bayes and structural improvement (Semi-Naïve Bayes)

The simple Bayesian classifier, named Naïve Bayes, is based on the Bayes rule expressed in (2.6.1) (Sucar, 2015):

$$P(C|A) = \frac{P(C)P(A|C)}{P(A)} \tag{2.6.1}$$

where C represents the class (hypothesis), and A the attributes (the evidence), which are generally expressed as: $A_1, A_2, ..., A_n$. Given the values of the attributes of an example $e_u = (a_1, a_2, \cdots, a_n)$, then the equation 2.6.1 can be expressed as in (2.6.2).

$$P(C = c|A_1 = a_1, A_2 = a_2, ..., A_n = a_n) = \frac{P(C = c)P(A_1 = a_1, A_2 = a_2, ..., A_n = a_n|C = c)}{P(A_1 = a_1, A_2 = a_2, ..., A_n = a_n)} \tag{2.6.2}$$

Estimate the probability $P(A_1 = a_1, A_2 = a_2, ..., A_n = a_n|C = c)$ is usually complex (Sucar, 2015), so this method, naively, assumes that the attributes or evidences are conditionally independent, among them, given the class (Mitchell, 1997). Thus, the simple Bayesian classifier states that the probability of the class given the attributes values can be calculated as indicated in (2.6.3).

$$P(C = c|A_1 = a_1, A_2 = a_2, ..., A_n = a_n) = \frac{P(C = c)\prod_{i=1}^{n} P(A_i = a_i|C = c)}{P(A_1 = a_1, A_2 = a_2, ..., A_n = a_n)} \tag{2.6.3}$$

Given $P(C = c)$ and $P(A_i = a_i|C = c)$ $i \in \{1, 2, ..., n\}$, it is not required to calculate the denominator $P(A_1 = a_1, A_2 = a_2, ..., A_n = a_n)$ because it is a fixed value for the different values of C (the different classes: c), i.e. $P(C = c|A_1 = a_1, A_2 = a_2, ..., A_n = a_n) \propto P(C = c)\prod_{i=1}^{n} P(A_i = a_i|C = c)$. The numerator provides the necessary information to discriminate the class c which is most likely. Classification consists in determining the value of C (class c) that maximizes expression (2.6.4).

$$class(e_u) = \arg\max_c (P(C = c)\prod_{i=1}^{n} P(A_i = a_i|C = c)) \tag{2.6.4}$$

where $P(C = c)$ is known as prior probability of the class, $P(A_i = a_i|C = c)$ $i \in \{1, 2, ..., n\}$ are the conditional probabilities of the evidences or attributes given the value of the class (likelihood), and $P(C = c|A_1 = a_1, A_2 = a_2, ..., A_n = a_n)$ is the posterior probability (the probability of the class).

This approach has the limitation that it is not always correct to assume the conditional independence of the attributes given the class (Mitchell, 1997; Sucar, 2015). Among the alternatives to correct this limitation, there exists the **structural improvement**, proposed as **Backward Sequential Elimination and Joining**: BSEJ by Pazzani (1996) and improved by Martínez-Arroyo and Sucar (2006). The algorithm of structural improvement uses the criterion of mutual information (MI) between each attribute A_i and the class C

(see 2.6.5), and the conditional mutual information between attribute pairs, given the class C (see 2.6.6) (Chow and Liu, 1968). Structural improvement uses MI to define whether the structure of the simple classifier should be modified, eliminating irrelevant attributes and/or joining dependent attributes (in pairs, combining their values) (see figure 2.2)(Pazzani, 1996; Martínez-Arroyo and Sucar, 2006; Martínez Arrollo, 2007).

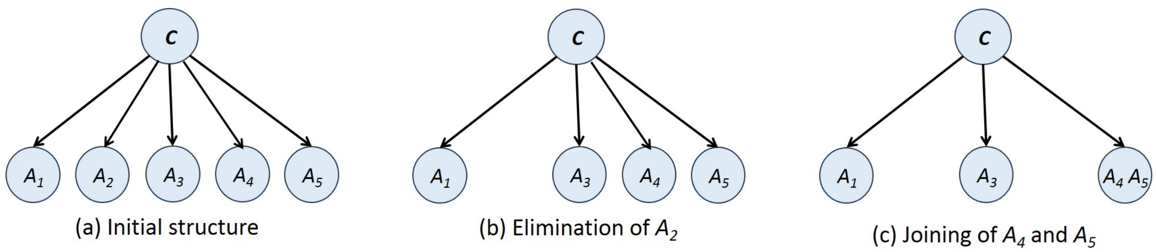


Fig. 2.2: Example of the structural improvement method: (a) example of the original structure of Naïve Bayes with 5 attributes (all attributes are assumed to be independent given the class), (b) attribute A_2 is removed, (c) attributes A_4 and A_5 are joined into one. This is an example of graphical model of a Semi-Naïve Bayes.

Mutual information between the attribute A_i and the class C (Chow and Liu, 1968) is calculated using (2.6.5).

$$MI\left(A_i, C\right) = \sum_{a_i \in A_i} \sum_{c \in C} P\left(A_i = a_i, C = c\right) \log \left(\frac{P\left(A_i = a_i, C = c\right)}{P\left(A_i = a_i\right) P\left(C = c\right)} \right) \tag{2.6.5}$$

The conditional mutual information between attributes A_i and A_j given the class C is obtained by (2.6.6).

$$MI\left(A_i, A_j | C\right) = \sum_{a_i \in A_i} \sum_{a_j \in A_j} \sum_{c \in C} P\left(A_i = a_i, A_j = a_j | C = c\right) \log \left(\frac{P\left(A_i = a_i, A_j = a_j | C = c\right)}{P\left(A_i = a_i | C = c\right) P\left(A_j = a_j | C = c\right)} \right) \tag{2.6.6}$$

2.6.2 Discretization method PKID

Continuous numeric attributes (attributes with values x $\in \mathbb{R}$, also known as ratio type attributes) must be discretized when using the Naïve Bayes classifier or the Semi-Naïve Bayes classifier. Proportional k-Interval Discretization (PKID) method proposed by Yang and Webb (2001) has been reported as a suitable alternative for these Bayesian classifiers since it seeks an appropriate compensation between bias and variability of the discretization intervals. In discretization methods there is the problem of defining the appropriate number of intervals (Yang and Webb, 2001); this number introduces a bias (as long as there are more intervals the bias decreases, in fact if the number of intervals coincides with the number of instances of the attribute, then there is no bias). The variability is associated with the number of instances in each interval, as long as more instances are available in the intervals, enough information is obtained to more accurately estimate the probabilities required by Bayes's Theorem (Yang and Webb, 2001).

The main idea of PKID is to calculate a number k of discretization intervals, so that the number of instances of the attribute, for each interval, is, as far as possible, k values; i.e., to match the number of intervals with the number of values that are placed in each interval. For this reason, if the attribute has N values, then $k = \lfloor\sqrt{N}\rfloor$.

2.6.3 Techniques of fusion in multimodal classification

The heterogeneity and variability of the data of some applications hinders a good performance when using only one classifier; this promotes the use of various classifiers and techniques to merge the results, with the intention of improving the performance of the classification (Mangai et al., 2010). An example of these applications is in the field of affective computing where data from different physiological sensors can be recorded (EEG, electrocardiogram (ECG), galvanic skin response (GSR), electromyogram (EMG), respiration and skin temperature (TEMP), among others) and it is desired to recognize the affective state of the person.

There are mainly three types of fusion strategies (Mangai et al., 2010):

1.- *data fusion* or information fusion (low-level fusion),

2.- *feature level fusion*, also known as *early fusion* (medium-level fusion), y

3.- *decision level fusion*, also known as *late fusion* (high-level fusion).

Data fusion: Original (raw) data from different sources are combined to obtain a new set of raw data, with the expectation that the new set will be more informative and will contain less noise than isolated sets of data from different sources (Mangai et al., 2010). The data to be merged must be on the same scale and meaning (Hu, 2008).

Feature level fusion (early fusion): Attributes generated from the original data are selected and/or combined to remove redundant attributes (attributes that have a similar distribution or close to others) or irrelevant ones (those that have little correlation with the class), and finally they are merged to produce a better set of attributes that will be the inputs to a classifier (Mangai et al., 2010). There are three general ways to merge the attributes (Hu, 2008): concatenation (attribute vectors of all sources are put together side by side to form a super attribute vector), selection (select a subset of attributes, among all available ones, based on some criterion), and extraction (all attributes available are projected to a new space).

Decision level fusion (late fusion): A set of classifiers is used to improve classification. An initial approach consists in using as many classifiers as sensors (one classifier per sensor, i.e., per each feature vector). The classifiers may be of the same or different type. When the feature vector resulting from an early fusion (by concatenation, selection, or extraction) is supplied, we have a hybrid fusion scheme (fusion at the feature level and fusion at the decision level). The outputs of all classifiers are combined using several methods to obtain the definitive output (Mangai et al., 2010).

Formally, each feature vector $\vec{x}_h$ (obtained from each data sensor) is provided to a classifier O_h, $h \in \{1, 2, ..., m\}$ which must infer the class corresponding to the vector $\vec{x}_h$ from the set of class labels $C = \{c_1, c_2, ..., c_d\}$. Each classifier O_h produces d degrees in the interval $[0, 1]$ represented by $o_{h,j}$ which indicate the level of support of O_h to the label c_j (of class variable C, $j \in \{1, 2, ..., d\}$). Class c_j that obtains the highest level of support will be the class assigned to $\vec{x}_h$ by O_h. Outputs of the h classifiers are organized in the matrix presented in 2.6.7 (Mangai et al., 2010).

$$\begin{array}{cc} & \begin{array}{ccccc} c_1 & \cdots & c_j & \cdots & c_d \end{array} \\ \begin{array}{c} 1 \\ \vdots \\ h \\ \vdots \\ m \end{array} & \begin{bmatrix} o_{1,1}(\vec{x}_1) & \cdots & o_{1,j}(\vec{x}_1) & \cdots & o_{1,d}(\vec{x}_1) \\ \vdots & & \vdots & & \vdots \\ o_{h,1}(\vec{x}_h) & \cdots & o_{h,j}(\vec{x}_h) & \cdots & o_{h,d}(\vec{x}_h) \\ \vdots & & \vdots & & \vdots \\ o_{m,1}(\vec{x}_m) & \cdots & o_{m,j}(\vec{x}_m) & \cdots & o_{m,d}(\vec{x}_m) \end{bmatrix} \\ & \begin{array}{ccccc} \mu_1(\vec{x}) & \cdots & \mu_j(\vec{x}) & \cdots & \mu_d(\vec{x}) \end{array} \end{array} \tag{2.6.7}$$

where $\vec{x} = (\vec{x}_1, \cdots, \vec{x}_h, \cdots, \vec{x}_m)$ and $\mu_j(\vec{x})$ represents the calculation per column (per class c_j) which depends on the method of fusion, among the methods: majority voting, sum rule, product rule, minimum rule and maximum rule.

Sum rule:

$$\mu_j(\vec{x}) = \sum_{h=1}^{m} o_{h,j}(\vec{x}_h), \qquad j \in \{1, 2, ..., d\} \tag{2.6.8}$$

Majority voting:
This is the particular case of the sum rule when each classifier O_h produces results $o_{h,j}(\vec{x}_h) \in \{0, 1\}$, so that $o_{h,j}(\vec{x}_h) = 1$ means that the classifier O_h predicts class c_j for the feature vector $\vec{x}_h$, and 0 otherwise. $\mu_j(\vec{x})$ is calculated according to 2.6.8.

Product rule:

$$\mu_j(\vec{x}) = \prod_{h=1}^{m} o_{h,j}(\vec{x}_h), \qquad j \in \{1, 2, ..., d\} \tag{2.6.9}$$

Min. rule:

$$\mu_j(\vec{x}) = \min_{h=1}^{m} o_{h,j}(\vec{x}_h), \qquad j \in \{1, 2, ..., d\} \tag{2.6.10}$$

Max. rule:

$$\mu_j(\vec{x}) = \max_{h=1}^{m} o_{h,j}(\vec{x}_h), \qquad j \in \{1, 2, ..., d\} \tag{2.6.11}$$

Definitive class:
Once the $\mu_j(\vec{x})$, $j \in \{1, 2, ..., d\}$ are generated, based on the respective equation (2.6.8,

2.6.9, 2.6.10 or 2.6.11) according to the rule applied; the definitive class c_j is calculated using 2.6.12.

$$clase(\vec{x}) = \underset{c_j, j \in \{1,2,\ldots,d\}}{\arg\max} (\mu_j(\vec{x})) \tag{2.6.12}$$

Another alternative to make the late fusion consist in using a classifier (SVM, artificial neural networks, *Naïve Bayes*, etc.) that receives as inputs the classes c_j inferred by the classifiers O_h for the respective feature vector $\vec{x}_h$.

2.6.4 Multidimensional classification

Several classification problems need to assign more than one class simultaneously to the feature vector $\vec{x}_h$. For example: in affective computing, where a photo can evoke a mixture of emotions instead of just one emotion. This type of classification where the objects can be tagged with several simultaneous classes is called multidimensional classification (Sucar, 2015).

Formally: Given the objects $u = (\vec{x}_u, \vec{c}_u)$, where the vector $\vec{x}_u = (x_1, x_2, \cdots, x_n)_u$, $x_i \in \Omega_{X_i}$, $i \in \{1, 2, ..., n\}$ is the feature vector, and $\vec{c}_u = (c_1, c_2, \cdots, c_d)_u$, $c_j \in \Omega_{C_j}$, $j \in \{1, 2, ..., d\}$ is the vector of class values assigned to $\vec{x}_u$. Given the parameters $\alpha \in \Lambda$, the goal consist in finding, within the set of admissible functions $f(\vec{x}_u, \alpha)$, the function $f(\vec{x}_u, \alpha_*) = \vec{c}_u$ $\forall u$ that assigns the most likely combination of classes values to $\vec{x}_u$ (that minimizes misclassification), as represented in 2.6.13.

$$f(\vec{x}_u, \alpha_*) = \underset{(c_1,c_2,\cdots,c_d)_u}{\arg\max} (P(C_1 = c_1, C_2 = c_2, \cdots, C_d = c_d|\vec{x}_u)) \tag{2.6.13}$$

where $f(\vec{x}_u, \alpha_*) : \Omega_{X_1} \times \Omega_{X_2} \times \cdots \times \Omega_{X_n} \rightarrow \Omega_{C_1} \times \Omega_{C_2} \times \cdots \times \Omega_{C_d}$

$$(x_1, x_2, \cdots, x_n)_u \mapsto (c_1, c_2, \cdots, c_d)_u$$

It is assumed that X_i $\forall i \in \{1, 2, ..., n\}$ y C_j $\forall j \in \{1, 2, ..., d\}$ are discrete, and Ω_{X_i} y Ω_{C_j} denote their sample spaces respectively.

Metrics (example based) to evaluate the performance of multidimensional classifiers

Several metrics have been established to evaluate the performance of multidimensional classifiers. Some of them are example based metrics: Given p: the number of examples in the data set; $c'_{u,j}$: the value predicted by the multidimensional classifier for the class j of the example u; $c_{u,j}$: the true value of the class j of the example u; $\vec{c}'_u$: the vector of values of the classes predicted by the multidimensional classifier for the example u; $\vec{c}_u$: the vector of true values of the classes for example u.

Exact match ratio (EMR) or Global accuracy (GAcc) (Sorower, 2010; Bielza et al., 2011) represents the extension of accuracy that is used in the traditional classification of a single class. GAcc is the accuracy for example.

$$GAcc = \frac{1}{p}\sum_{u=1}^{p}\bigwedge_{j=1}^{d}(c'_{u,j} = c_{u,j}) \tag{2.6.14}$$

where the result of the operator $\bigwedge_{j=1}^{d}$ is 1 to indicate true in all the expressions depending on j and 0 to indicate false in any of the expressions depending on j.

This metric is quite strict and does not consider whether the classification has been partially correct in each example.

Mean accuracy (MAcc) (Bielza et al., 2011) represents the accuracy by class, in this case the results that are partially correct are taken into account.

$$MAcc = \frac{1}{d}\sum_{j=1}^{d} Acc_j = \frac{1}{d}\sum_{j=1}^{d}\frac{1}{p}\sum_{u=1}^{p}\delta(c'_{u,j}, c_{u,j}) \tag{2.6.15}$$

where Acc_j is the calculation of accuracy for the class j and $\delta(c'_{u,j}, c_{u,j}) = 1$ if $c'_{u,j} = c_{u,j}$ and 0 otherwise.

Multi-label accuracy (MLAcc), also called Jaccard measure (Godbole and Sarawagi, 2004; Sorower, 2010) is the proportion of predicted correct labels to the total number of labels (predicted and true) for that example, averaged over all examples.

$$MLAcc = \frac{1}{p}\sum_{u=1}^{p}\frac{|\vec{c}'_u \cap \vec{c}_u|}{|\vec{c}'_u \cup \vec{c}_u|} \tag{2.6.16}$$

where in $|\vec{c}'_u \cap \vec{c}_u|$ we count the number of coincidences of the two vectors (predicted and true) and in $|\vec{c}'_u \cup \vec{c}_u|$ we count the number of labels covered by those vectors.

Precision (P) (Godbole and Sarawagi, 2004; Sorower, 2010) is the proportion of predicted correct labels to the total number of labels inferred by the classifier, averaged over all examples.

$$P = \frac{1}{p}\sum_{u=1}^{p}\frac{|\vec{c}'_u \cap \vec{c}_u|}{|\vec{c}'_u|} \tag{2.6.17}$$

where in $|\vec{c}'_u|$ we count the number of labels of the vector $\vec{c}'_u$.

Recall (R) (Godbole and Sarawagi, 2004; Sorower, 2010) is the proportion of predicted correct labels to the total number of true labels, averaged over all examples.

$$R = \frac{1}{p}\sum_{u=1}^{p}\frac{|\vec{c}'_u \cap \vec{c}_u|}{|\vec{c}_u|} \tag{2.6.18}$$

F-measure (Godbole and Sarawagi, 2004; Sorower, 2010) is the harmonic mean between precision and recall.

$$F - measure = \frac{1}{p}\sum_{u=1}^{p}\frac{2\,|\vec{c}_u' \cap \vec{c}_u|}{|\vec{c}_u'| + |\vec{c}_u|} \tag{2.6.19}$$

Bayesian chain classifiers

An alternative method for multidimensional classification consists in using chain classifiers, whose architecture allows the incorporation of dependencies among classes and maintains the computational efficiency of binary relevance method (Read et al., 2009) (binary classifiers are proposed to infer each class separately). If class vector is $\vec{c} = (c_1, c_2, \cdots, c_d)$, then a chain of d base binary classifiers (one per class) are required and linked so that each classifier incorporates among its input features the values of the predicted classes (0/1 or $-1/1$) by the previous classifiers in the chain. Thus, the feature vector for each binary classifier O_h, $h \in \{1, 2, ..., d\}$ extends with an additional label (0/1 or $-1/1$) for each previous classifier $O_1, O_2, \cdots, O_{h-1}$, in the chain. Each classifier O_h in the chain is trained to learn the association between the feature vector $(\vec{x}, c_1', c_2', \cdots, c_{h-1}')$ and the class label c_h; where $\vec{x}$ is the original feature vector and $c_1', c_2', \cdots, c_{h-1}'$, are the classes calculated by the previous classifiers. In the testing phase, the classification starts at O_1 and the predicted classes are propagated along the chain, in a way that for O_h it is inferred $P(c_h|\vec{x}, c_1', c_2', \cdots, c_{h-1}')$. At the end, the vector of the predicted classes is formed from the classes obtained by each classifier O_h (Sucar, 2015).

Bayesian chain classifiers are a particular type of chain classifiers that arise when the chain rule of probability theory is applied and the chain is constructed considering the precedences based on the parent classes obtained from the dependency relations (based upon the structure of an directed acyclic graph (DAG)) between classes. The base classifier can be Naïve Bayes or any other Bayesian classifier.

2.7 Summary

The definitions of emotion and affective state have been established. It should be noted that the term affective state instead of emotion will be used in the next chapters to have the possibility of including other states that are not exactly an emotion (and avoid controversies with experts in the field of psychology). The Personality Big Five Factor Model will be used as a tool to identify the personality of the participants at the data collection. The central learning method of this research is the paradigm of learning using privileged information; on the other hand, the multimodal classification using Bayesian chain classifiers is important for this proposal for its computational efficiency and to take advantage of the relations of dependence of the affective states.

Chapter 3

Related works

For addressing this research work, it was necessary to investigate the advances achieve so far in sub-areas such as: application of affective computing in computer games, development of multimodal systems for the recognition of affective states, and finally, learning using privileged information and multidimensional classification in affective computing.

3.1 Affective videogames

In computer game development, affective computing has been incorporated to try to improve the user experience. Among the affective states that have been studied in this area are: boredom, engagement (interest), anxiety, frustration, excitement, joy (rejoicing), calmness (relaxation), anger (irritation) surprise, sadness and fatigue (Chanel et al., 2011; Gilleade and Dix, 2004; Chumbley and Griffiths, 2006; Van Reekum et al., 2004; Bianchi-Berthouze et al., 2007; Savva and Bianchi-Berthouze, 2011; Gao et al., 2012; Rivas, 2015), among others. One of the benefits consist in adapting the game according to the user's specific needs. The purpose should be to maintain a state of pleasure to the player (Chanel et al., 2011). This is particularly relevant in serious games, to stimulate the player to become more interested in the assigned task (Shaker et al., 2010; Gilleade and Allanson, 2003; Conati, 2002; Molins-Ruano et al., 2014; Bonarini et al., 2010). The affective states can be induced by changing the difficulty level of the game but this must be adjusted in concordance with the player skills (Chanel et al., 2011; Gilleade and Dix, 2004; Chumbley and Griffiths, 2006; Van Reekum et al., 2004).

3.2 Multimodal recognition of affective states

The recognition of emotions in affective computing took an important turn when several authors began to argue the need to incorporate several sources of user data to achieve more reliable results (Chen et al., 1998; Picard et al., 2001; Pantic et al., 2007; Cernea and Kerren, 2015). Accordingly, there have been studies combining facial expressions with physiological signals (Bailenson et al., 2008), facial expressions with speech (Zhalehpour et al., 2014), physiological signals from multiple sources (Picard et al., 2001; Calvo et al., 2009), speech and physiological signals from multiple sources (Kim et al., 2009), EEG and visual tracking (Zheng et al., 2014), EEG and peripheral physiological signals from multiple sources (Chanel et al., 2011), achieving better results than those obtained with a single data source. The table 3.1 presents a summary of these works indicating the multimodal elements, how the affective states were induced, the specific affective states studied, the classifiers, the fusion strategy and the results.

Table 3.1: Summary of some emotion recognition works with multimodal approach.

Work	Multimodal elements	Stimulus	Affective states	Classifier	Fusion	Results
(Bailenson et al., 2008)	facial expressions and physiological signals (electrocardiogram (ECG), galvanic skin response (GSR))	video clips with emotional content	sadness and enjoyment	SVM with linear kernel	feature concatenation (early fusion)	F-measure: Per person (average): sadness = 0.95 and enjoyment = 0.92. All: sadness = 0.37 and enjoyment = 0.66
(Zhalehpour et al., 2014)	facial expressions and speech	emotions were acted	basic discrete emotions: anger, disgust, fear, happiness, sadness and surprise	SVM (linear kernel) for facial expressions and SVM (radial base kernel) for speech	at decision level (late fusion): product rule	accuracy: 76.4 %
(Calvo et al., 2009)	Physiological signals (electromyogram (EMG), ECG and GSR)	self-induced affective states through meditation using the protocol *Sentics Clynes* (Clynes and Menuhin, 1977)	no emotion, anger, hate, grief (pain), platonic love, romantic love, joy and reverence	comparison of 8 classifiers: ZeroR, OneR, function trees, Naïve Bayes, Bayesian networks, multilayer Perceptron, linear logistic regression and SVM	feature concatenation	chosen classifier: SVM. Accuracy: 3 people: 1(80.4%), 2(85.7%) and 3(79%). All(42.2%)
(Kim et al., 2009)	Physiological signals (EMG, ECG, GSR, respiration changes) and speech	emotions induced by a set of questions that evoke situations	discrete arousal/valence: high/positive, high/negative, low/positive and low/negative	pseudoinverse of Linear Discriminant Analysis (pLDA) and Sequential Backward Selection (SBS)	three approaches: 1-feature concatenation, 2-majority vote in decision level and 3-hybrid: majority vote including result of 1	highest accuracy was achieve using fusion approach 1: 3 people: A(92%), B(75%) and C(69%). All(55%)
(Zheng et al., 2014)	EEG and eye tracking	video clips with emotional content	discrete valence (positive, neutral and negative)	SVM	three: 1-feature concatenation, 2-decision level: max rule and 3-decision level: sum rule	highest accuracy was obtained using fusion approach 1: $73.59\% \pm 14.43$
(Chanel et al., 2011)	EEG and peripheral physiological signals (GSR, respiratory rate, blood volume pulse (BVP) and hand temperature)	emotions induced through interaction with a computer game	boredom, engagement and anxiety	Linear Discriminant Analysis (LDA) for EEG signals and Quadratic Discriminant Analysis (QDA) for physiological signals	at decision level: integration by Bayes belief integration	accuracy: 63%

None of the works described in table 3.1 considers the problem of using the system in everyday life and the obstructive difficulties of using the sensors. Only the work of (Zhalehpour et al., 2014) could be used outside the laboratory, however, facial expressions and speech do not always reflect the affective states that the user is really experiencing (Wang et al., 2015d; Zheng et al., 2014).

3.3 Multimodal data set for studying affective states

Two multimodal data repositories (Vinola and Vimaladevi, 2015; Soleymani and Pantic, 2012), MAHNOB-HCI and DEAP are available to the academic community.

MAHNOB-HCI (*Multimodal Analysis of Human Nonverbal Behavior)* (Soleymani et al., 2012): It contains the signals of EEG (32 channels), eye gaze tracking and four peripheral physiological signals: electrocardiogram (ECG), galvanic skin response (GSR), respiration amplitude (RSP) and skin temperature (TEMP), of 27 participants who watched 20 video clips with emotional content (to induce 5 emotions: disgust, amusement, joy, fear and sadness; and a neutral state) during the capture of the signals. Participants answered a questionnaire at the end of each video, to estimate in the discrete scale from 1 to 9 (1 = lowest level) the values of valence and arousal of the emotion they experienced. Correspondence was also made with discrete emotions: neutral, anxiety, amusement, sadness, joy, disgust, anger, surprise, and fear

DEAP (*Database for Emotion Analysis using Physiological signals*) (Koelstra et al., 2012): It stores EEG responses and peripheral physiological signals (GSR, RSP, TEMP, ECG, blood volume by plethysmograph (PLET), electromyogram (EMG) and electrooculogram (EOG)) to the stimuli of 40 video clips, to which 32 participants were exposed. Additionally it contains the front video of the face of 22 of the participants. As in the MAHNOB-HCI database, the participants responded to a self-assessment, estimating the valence, arousal and dominance levels of the emotion in the discrete scale from 1 to 9 (1 = lowest level).

These databases will be very useful to compare our proposal with the closest related works, which are described in the following section.

3.4 Learning using privileged information

The learning using privileged information paradigm was first introduced by Vladimir Vapnik and Akshay Vashist (Vapnik and Vashist, 2009) very oriented towards implementation in the field of SVM; The implementation in other machine learning techniques is under investigation (Wang et al., 2014b; Wang and Ji, 2015). Specifically, in the field of affective computing, the potential of this paradigm has been considered because of the possibility that during the training phase, several sensors will be available and some of them will register privileged information, which will not be easily acquired during the test phase. One of the applications has been the implicit tagging of emotional videos, in which the observer's physiological responses are analyzed to label the segments of the video with the different emotions that the user has experienced (Soleymani and Pantic, 2012). In Wang et al. (2015d), the implicit video emotion tagging was studied, and also the recognition of emotions using EEG signals, and the information of videos used to induce emotions. Through canonical correlation analysis (CCA) two new feature spaces were created, one for the EEG and the other for the video, which encapsulated the relationship between the features of both sources (EEG and video); two SVMs were trained respectively over each feature space. The SVM built over video space uses the EEG features as privileged information and serves to make the implicit video tagging; and the SVM from the EEG feature space, uses the features of the video as privileged information to recognize emotions. Experiments were performed on the MAHNOB-HCI, DEAP and USTC-ERVS databases (USTC-ERVS (Wang et al., 2014a) is not publicly available) and the valence and arousal classifications were better than the classical SVM classifications. The drawback of this approach is that it does not provide a solution for the case of more than two sensors. Chen et al. (2016) developed an SVM model with similarity restrictions in the mapping functions to capture the relationship between EEG signals, multiple user peripheral physiological signals (EOG, EMG, ECG, GSR, RSP, TEMP and PLET) and the features of video content. In this case the EEG signals and the different peripheral physiological signals represent the privileged information in the implicit video tagging, i.e. during the test phase (video tagging) only the video features are available. The authors also used the MAHNOB-HCI, DEAP and USTC-ERVS databases for testing. Instead of using only one physiological signal as (Wang et al., 2015d), its approach can integrate multiple physiological signals to facilitate the video emotion tagging. These results are used for implicit video tagging applications, but in other applications how it would work, for example in environments of emotions induced by interaction with video games.

Learning using privileged information has also been applied to model individual differences and general patterns in the EEG signals of various subjects for automatic emotion recognition (Wu et al., 2016). This new approach allows to use the individual information of each subject as privileged information, or use the general information of the subject group as privileged information, which is only available during training. Two Bayesian network structures were tested on the MAHNOB-HCI, DEAP and USTC-ERVS databases to predict the valence label (positive or negative) or arousal label (high or low); and through the joint probability distribution learned by the Bayesian networks, the tags could be estimated from the features of EEG only, marginalizing on the privileged information: the subject or the group of subjects. This approach was also studied using hierarchical Bayesian networks to handle the generality and specificity simultaneously of the EEG signals of the group of individuals in automatic recognition of emotions (Gao and Wang, 2015).

Bayesian networks inherently allow marginalization of features and, for this reason, are useful for learning using privileged information (Wang et al., 2015a). Basically, Wang et al. (2015a) studied Bayesian structures of three general nodes: the class node y (the emotion variable), the available information node x and the privileged information node x^*, and all possible connections and directions of the arcs between these nodes.

3.5 Multidimensional classification including dependence relationships among emotions

Very few works consider dependence relationships among emotions and multidimensional classification (Wang et al., 2015b). The works of Wang et al. (2013) and Wang et al. (2015b) exploit these characteristics. In the first work, a Bayesian network is used to learn the relations of co-occurrence and mutual exclusion among pairs of emotions; this is somewhat limiting because it do not include dependency relationships among more than two emotions simultaneously. In the second, a three-layer Boltzmann restrictive machine is used to identify the relations of dependence among more than one pair of emotions; The problem is the computational cost involved in training and making inferences in the Boltzmann machine. In our proposal, we consider dependency relations of two or more affective states by adding the previous classes in the chain sequence of Bayesian classifiers, in addition it is proposed to use MSNB (Multiresolution Semi-Naïve Bayes) classifier, whose base classifiers are Semi-Naïve Bayes that maintain the efficiency and simplicity of the Naïve Bayes classifiers (Sucar, 2015).

3.6 Summary of the closest related works to our proposal

Our research work is based on learning using privileged information and multidimensional classification in the area of affective computing, using computer games environments. In the Table 3.2, the characteristics of the related works and their proximity to the present investigation are shown; indicating the multimodal elements, the stimulus used to induce the affective states, the affective states studied, the privileged information, the classification strategies developed and the results obtained. The last row of the table describes the components of our proposal. The process of comparing the present work with related works will be achieved using the MAHNOB-HCI and DEAP databases, and making the classification of discrete valence and discrete arousal under the same conditions.

The works of Wang et al. (2015d) and Chen et al. (2016) use SVM as the fundamental classifier; but in the previous work of Rivas (2015) SVM was compared with the Semi-Naïve Bayesian classifier and with the Multiresolution Semi-Naïve Bayesian classifier, and both classifiers achieved better results than SVM on accuracy, sensitivity, specificity , precision, F-measure and ROC area for the automatic recognition of tiredness, anxiety, pain and engagement of two participants. On the other hand, novel elements of our proposal are the inclusion of hand movements and fingers' pressure to recognize affective states, and to apply learning using privileged information in videogames environments.

Table 3.2: Summary of related works for recognition of affective states through learning using privileged information.

Work	Multimodal elements	Stimulus	Affective states	Privileged information	Classification strategy	Valence results	Arousal results
(Wang et al., 2015d)	EEG and content features of videos which induce emotional experiences	video clips which induce emotional experiences	discrete valence: {positive, negative}; discrete arousal: {high, low}	EEG or video content	canonical correlation analysis (CCA) and SVM	accuracy: For MAHNOB-HCI: Video as privileged info: 58.16%, EEG as privileged info: 60.22%. For DEAP: Video as privileged info: 59.70%, EEG as privileged info: 59.46%	accuracy: For MAHNOB-HCI: Video as privileged info: 61.35%, EEG as privileged info: 61.35%. For DEAP: Video as privileged info: 56.91%, EEG as privileged info: 55.59%
(Chen et al., 2016)	EEG, content features of videos which induce emotional experiences and peripheral physiological signals (EOG, EMG, ECG, GSR, RSP, TEMP and PLET)	video clips which induce emotional experiences	discrete valence: {positive, negative}; discrete arousal: {high, low}	EEG or the peripheral physiological signals	modified SVM with similarity restrictions in the mapping functions	accuracy: For MAHNOB-HCI: EEG as privileged info: 75.23%, peripheral signals as privileged info: 75.23%. For DEAP: EEG as privileged info: 71.05%, peripheral signals as privileged info: 73.68%	accuracy: For MAHNOB-HCI: EEG as privileged info: 84.99%, peripheral signals as privileged info: 84.99%. For DEAP: EEG as privileged info: 78.95%, peripheral signals as privileged info: 78.95%
our proposal	EEG, peripheral physiological signals (EOG, EMG, ECG, GSR), eye tracking, hand movements and fingers' pressure	affective states induced through the interaction with a computer game	discrete affective states: {tiredness, anxiety, boredom, engagement}	mainly: EEG, peripheral physiological signals or eye tracking	base classifier: Multiresolution Semi-Naïve Bayes. Apply Semi-Naïve Bayes or a Bayesian network at decision level. In addition, the dependence relationships among affective states will be exploited through Bayesian chain classifiers	-	-

Chapter 4
Methodology

4.1 Considered challenges

To achieve the aim of developing and validating a multimodal computational model of recognition of affective states that allow the marginalization of some sensor, and taking advantage of the relations of dependence of the affective states in the multidimensional classification; there are aspects of some of affective computing challenges that will be addressed (see section 1.1). The following are specific proposals for some of these challenges:

1.- The management of uncertainty will be done with the use of probabilistic graphical models.

2.- This research includes the development of multimodal models with physiological signals and external behavioural information.

3.- The context in which affective states are developed will be subtly considered by registering the personality and gender of participants during the construction of the own data set, and determining how these information can influence the results.

4.- The feasibility of using the system at home and/or in everyday places could be achieved through the idea of using marginalization of some of the sensors.

4.2 Multiresolution Semi-Naïve Bayesian classifier (MSNB)

One difficulty for detecting emotions is their sudden appearance (generated by some stimulus) and the fact that their duration is highly variable (Verduyn et al., 2015); in some cases relatively short (Scherer, 2005). No consensus has been reached on how long the emotions last (Scherer, 2005; Cabanac, 2002). In a previous work (Rivas et al., 2016) a binary classifier was proposed to explore the appearance of affective states of interest in the trace over time. The classifier operationalizes several odd-size windows (starting from 3) concentric to a current point that shift simultaneously over the trace, and which it becomes possible to calculate, in the current point environment, several features that may help to discriminate the presence of the affective state (see figure 4.1). This classifier was called Multiresolution Semi-Naïve Bayesian classifier (MSNB) because the windows represent several simultaneous resolutions at the current point of the trace. The classifier represents an ensemble of Semi-Naïve Bayesian classifiers (SNB) with a late (decision level) fusion process by majority vote. Each SNB receives the features coming from a different window size and infers the presence or not of the affective state of interest. At the end, in the fusion stage, the presence or not is decided, by means of the majority vote of the SNBs. In part b) of the figure 4.1 the architecture of MSNB is shown. Because of the good results obtained in the previous work (Rivas et al., 2016), it was decided to use MSNB as the base classifier for the computational model to this proposal.

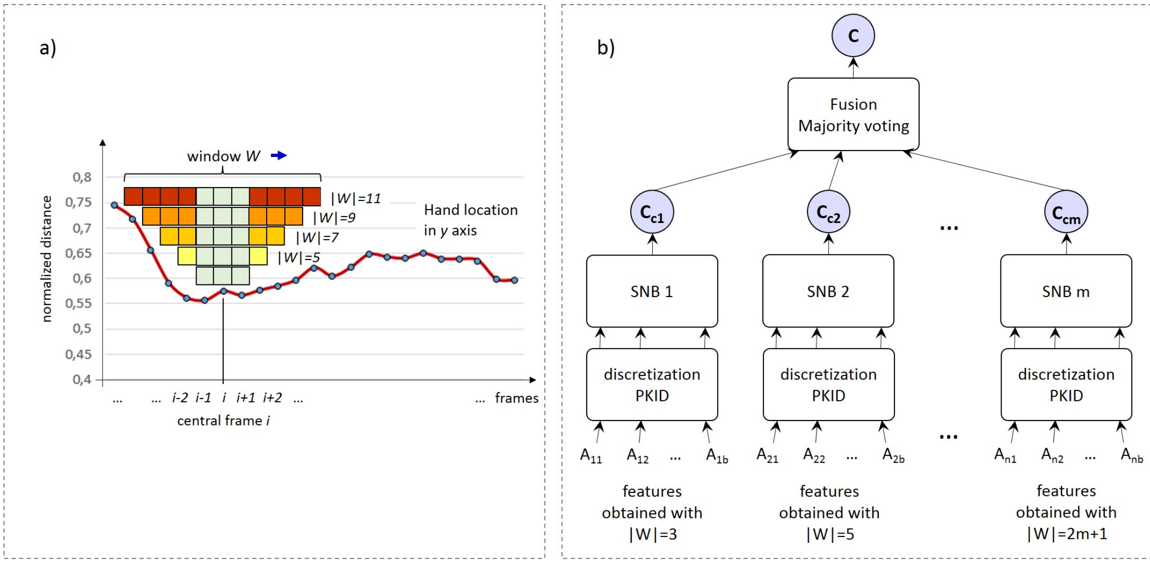


Fig. 4.1: Multiresolution Semi-Naïve Bayesian classifier (MSNB). In a), we present an example of the process of multiresolution with the use of windows, of different odd size, $|V| = 3{,}5{,}7{,}9{,}11$, for the trace of the hand in the axis y. This example, for reasons of simplicity and clarity, shows the trace in one of the axes; however it is necessary to visualize the concept with the 3D hand movement trace. For each window a semi-Naïve Bayesian model is constructed to estimate the presence or not of the affective state; Then by majority voting from the 5 models, it is inferred whether the affective state is present or not in the environment of the current point of the trace. In b) the architecture of the classifier is presented. Features from several concentric odd-size window $|V| = 2m + 1$ with respect to a point of the trace, are supplied and discretized with PKID method (if features are continuous). The discretized vectors constitute the inputs to the semi-Naive Bayesian classifiers which independently decide whether or not the affective state exists for that window size $|V|$. These inferences are received by the module of late fusion where, by majority voting, finally decides whether the affective state is present or not.

4.3 Methodological steps:

The methodological steps proposed to achieve the objectives are presented in the block diagram of the figure 4.2.

Step 1 Start by studying the first versions of the models of marginalization and multidimensional classification to determine the feasibility of obtaining competitive results with previous work (Rivas et al., 2016). This step is part of what was reported in the preliminary results.

(a) Choice of an affective data set containing signals for few sensors: two or three sensors. The data set of the previous work (Rivas et al., 2016) was chosen, which records the signals of two sensors: MOV (hand movements) and PRE (finger pressure) of two stroke patients; to recognize the state of tiredness, anxiety, pain and engagement. For more details see the section 5.1.

(b) Design and preliminary development of the multimodal computational model to address the marginalization of some of the sensors (MOV or PRE). In this

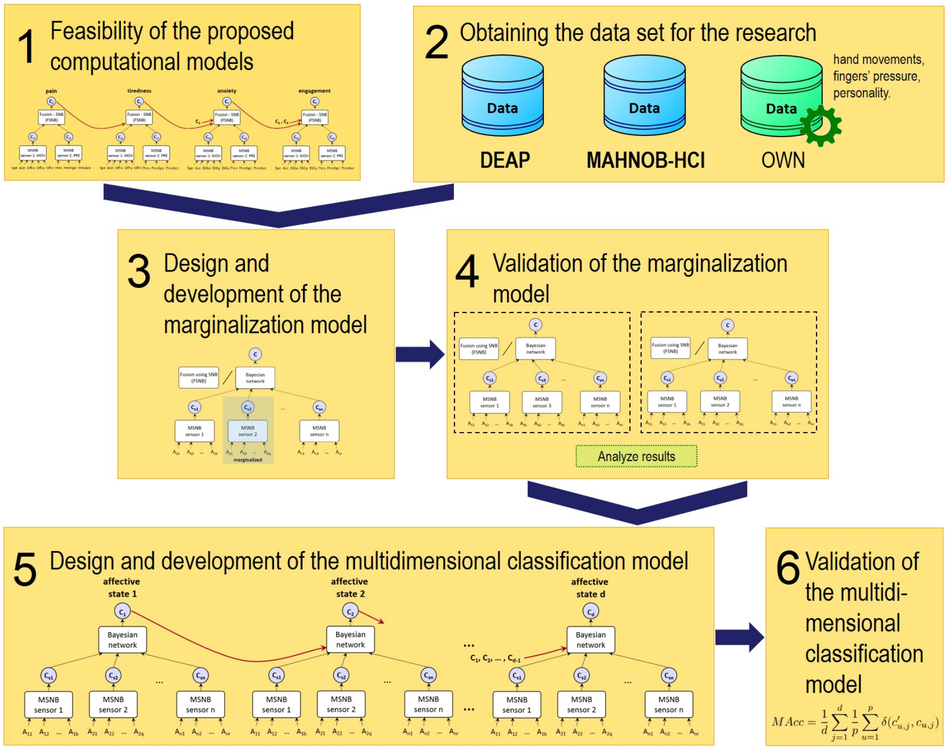


Fig. 4.2: Steps of the methodology.

step was considered as base classifier MSNB to infer the relationship among the features of each sensor and the affective state of interest. A MSNB is used for each sensor and the Semi-Naïve Bayesian (SNB) classifier is used in the late (decision level) fusion. The complete model is called fusion using SNB (FSNB). The FSNB architecture is presented in the figure 4.3.

(c) Experimentation and evaluation of the results of the proposed computational model, marginalizing one sensor at a time and determining the effectiveness of the selected marginalization strategy. There are three scenarios: (a) The model with the two sensors available, i.e. the complete model without marginalization, (b) the model where a sensor is marginalized; and (c) the model where only the available sensor exists, i.e. where there is no influence of the marginalized sensor. Accuracy, F-measure, and ROC area metrics were used to compare performance. The results of the three models were contrasted to see if the model (b) of marginalization achieve a better performance than the model (c).

(d) Initial development of the model for the multidimensional classification that incorporates the relations of dependence among the affective states. The model of Bayesian chain classifiers is proposed, with the novelty of linking the chain in two levels: sensors level and/or classes level (see subsection 4.3.3 where the architecture of this model appears).

(e) Experimentation and evaluation of multiclassification results. The results are compared with those obtained by the individual classifiers of the affective states. In this stage, it will be known if the multiclassification strategy took advantage of the relations of dependence of the affective states to improve the classification.

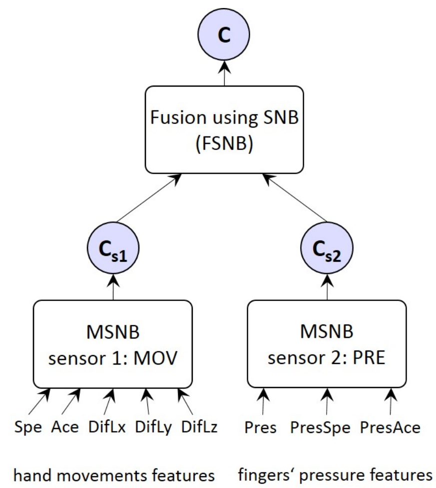


Fig. 4.3: Architecture of the fusion using SNB (FSNB). MSNB is used as the base classifier; independently it receives the features of MOV sensor and features of PRE. Finally the fusion is made at the decision level through SNB. Hand movements features are represented as: (average of) speed (Spe), acceleration (Ace), differential location x (DifLx), differential location y (DifLy) and differential location z (DifLz); fingers' features are indicated as: (average of) pressure (Pre), pressure speed (PresSpe) and pressure acceleration (PresAce).

Step 2 Getting the dataset for research. These datasets should contain physiological signals and external behavioural information; and the annotation of at least 3 discrete affective states (to study the dependency relations among them). The datasets will come from repositories published in the literature, as well as generated specifically for this thesis.

(a) From the datasets in the literature we will try to obtain DEAP (Koelstra et al., 2012) and MAHNOB-HCI (Soleymani et al., 2012) that were used in the related works (Wang et al., 2015d; Chen et al., 2016).

(b) In order to construct a multimodal data set of affective states, it is intended to capture experimental data that includes among the sensors, the recording of hand movements and fingers' pressure. The personality of each participant will also be tabulated. This set of data will allow us to study the contribution of hand movements and fingers' pressure to the recognition of affective states, in conjunction with the other sensors that were chosen. Additionally, trying to study the context, it is convenient to evaluate the results of the computational model contrasting with the personality and gender of the participants. In terms of personality registration, the *Big Five Personality Scale* model questionnaire will be used. (Renau et al., 2013; Romero et al., 2012; Goldberg, 1990; Costa and MacCrae, 1992; Tupes and Christal, 1961).

It should be emphasized that this data set will be an additional contribution of this thesis; however, with the DEAP and MANHOB-HCI datasets, the proposed model can already be evaluated and comparisons can be done with the related works; and consequently the necessary publications can be prepared.

Step 3 Design and development of the model of marginalization strategy.

(a) Start designing the architecture of the model that will integrate the signals of all sensors, considering the synchronization among them. This model requires fusion of the information; it is proposed to make fusion at the decision level (late fusion). In the subsection 4.3.1 the proposal is presented which consists of the base classifier MSNB for each sensor and the fusion at the decision level using a Bayesian network or using SNB.

(b) Incorporate into the designed architecture the mechanisms that achieve the marginalization scheme with the approach of learning using privileged information. The model must have the flexibility of considering any sensor as privileged information, i.e. during the testing phase any of them can be left aside. Bayesian networks allow the natural marginalization of some of its variables. The subsection 4.3.2 illustrates the marginalization of one of the sensors.

Step 4 Validation of the marginalization model.

(a) Build the experimental design to evaluate the computational model developed in the previous step. For the analysis, the confusion matrix metrics will be used in the binary classification: absence or presence of each affective state (accuracy, sensitivity, specificity, precision, F-measure and ROC area) and the respective models of the three cases: (i) case where all sensors are available, (ii) case of the model with marginalized sensors and (iii) case using only non-marginalized sensors; will be used.

(b) Adaptation to make comparisons of the proposed model with related works. The related works that will serve to compare our approach are Wang et al. (2015d) and Chen et al. (2016). The DEAP and MAHNOB-HCI data sets are used in both works.

Step 5 Design and development of the model for multidimensional classification.

(a) Study and determine from the literature, with the advice of experts in psychology or psychiatry, the dependence that may exist among the affective states considered. Outline the precedence, from the less dependent affective state to the one with the most dependent relationships with others.

(b) Design the model for the multidimensional classification that takes advantages of the dependency among affective states. At this point the proposal consist in using Bayesian chain classifiers (Sucar et al., 2014). In the subsection 4.3.3 the architecture model is presented according to this approach, with the contribution of a novel 2-level system.

(c) Construct the multidimensional classification model. The idea is to use the classifier proposed in step 3 and link the classifiers in a chain at the level of MSNB (classes calculated by the sensors) or at the level of the final class obtained after fusing by Bayesian networks or by SNB.

Step 6 Validation of the computational model for multidimensional classification.

(a) Construct the corresponding experimental design, taking into consideration the possibility of studying other sequences in the chain of affective states.

(b) Develop the experiments and compare the results of the multidimensional classification with the results of the individual classification of each affective state. Use metrics of the binary classification confusion matrix and multidimensional classification metrics, such as: global accuracy, mean accuracy on the number of affective states, multi-label accuracy, accuracy, sensitivity and average F-measure on the number of affective states.

4.3.1 Architecture of the proposed model for the integration of sensors

In the figure 4.4 the proposed architecture is displayed to integrate the information of the signals of the sensors. The features of the signals of each sensor are processed by MSNB, which constitutes the base classifier. Each MSNB infers the presence or not of the affective state of interest from the respective sensor that corresponds to it. The classes of each MSNB are received in the fusion phase. A Bayesian network integrates the values of the variables (classes inferred by the MSNB) and performs the total abduction to infer the most probable value of the class (absence or presence of the affective state). When the fusion is done by SNB, the classifier infers with the structure obtained after structural improvement, whether the affective state is present or not.

4.3.2 Architecture of the model proposed for marginalization

A sensor provides the information of a set of variables to the model. When one of the sensors is not available, the marginal distribution of its observations must be inferred from the rest of the available information (the variables of the available sensors). The process of marginalization is illustrated in figure 4.5, where it is exemplified using sensor 2 as marginalized. The information learned during training (when the values of sensor 2 were available) it is used to determine the presence or absence of the affective state. Since the classifier used in the last level of the model is a Bayesian network, then a partial abduction process is performed, i.e., the values of some variables of the 2nd level (classes inferred by the non-marginalized sensors) represent the evidence, and the most probable value of the class (presence or absence of the affective state) is calculated, identifying the variables whose

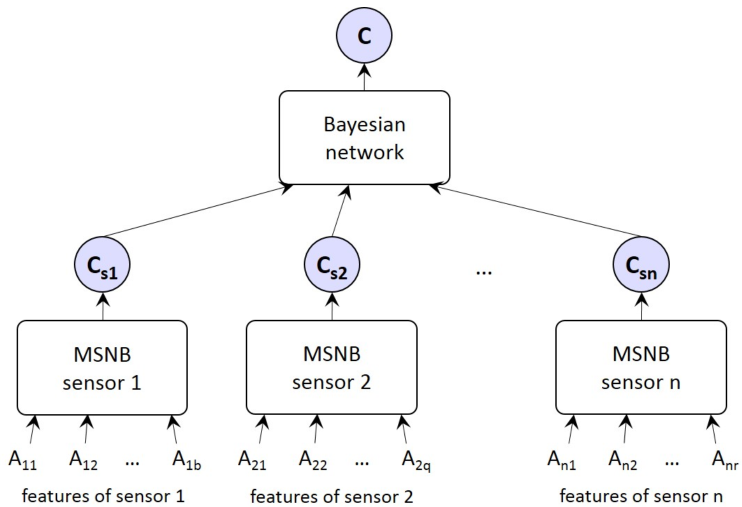


Fig. 4.4: Architecture of the fusion model at the decision level using a Bayesian network. For each sensor a MSNB classifier is used to estimate the presence or absence of the affective state under consideration. The fusion is made at the decision level (late fusion) using a Bayesian network that receives the values of the variables (classes inferred by the MSNB of each sensor) and through abduction decides which is the most probable value of the class (absence or presence of the affective state).

values are not present (marginalized sensors) as marginalized variables by the Bayesian network.

4.3.3 Architecture of the proposed model for multidimensional classification

For multidimensional classification it is proposed to use Bayesian chain classifiers. The Bayesian classifier to be used is the model that was developed in the previous subsection 4.3.1. As there are two levels, the first refers to the MSNB of each sensor, and the second, to the final integration through the Bayesian network that indicates the inferred value for the class; then the chain can be linked at the MSNB level (sensors level) and/or at the level of the classes. The chain at sensors level consists in linking the classes inferred for the MSNB of sensors of the same type in the chain. On the other hand, the classes obtained in the previous Bayesian networks (affective states) are incorporated as additional features in each Bayesian network, in fact, the number of variables of each Bayesian network is increasing. There are three possibilities to study this model, applying each chain separately or applying both chains at the same time. In the figure 4.6 the Bayesian chain classifier is illustrated.

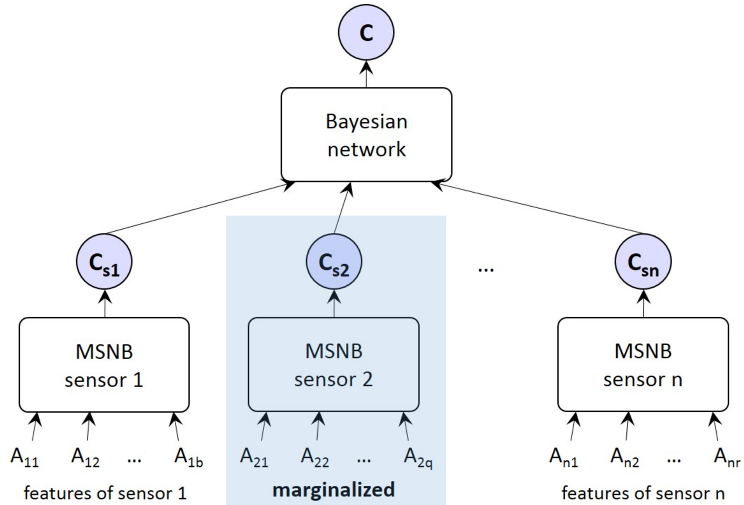


Fig. 4.5: Architecture of the integration of the signals through a Bayesian network which illustrates the marginalization of the sensor 2. The MSNB of the available sensors provide their information of presence or absence of the affective state. The Bayesian network as an integrating agent receives the information (classes) from the MSNBs of the available sensors (represents the evidence). For unavailable sensors, their respective MSNB can not give information, so the Bayesian network marginalizes them, and by partial abduction the network calculates the most probable value of the class, which indicates whether the affective state is present or not.

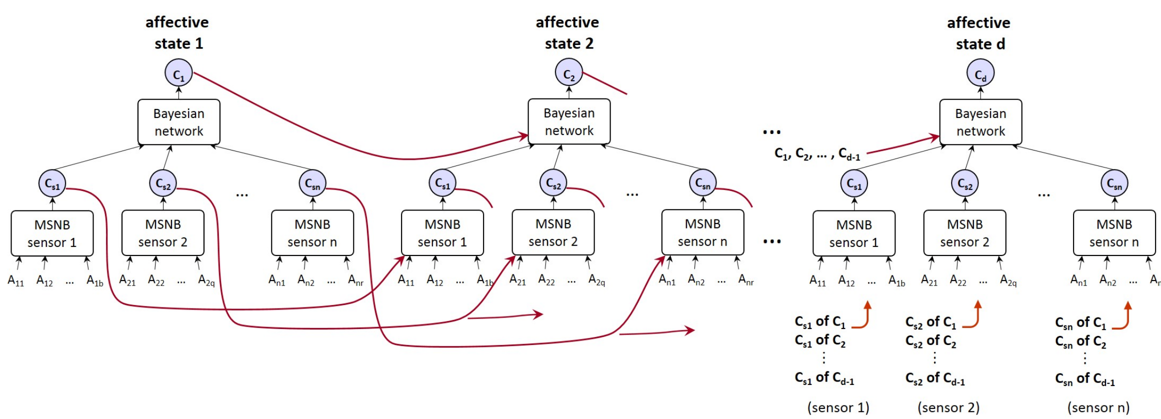


Fig. 4.6: Suggested architecture for the multidimensional classification model. Bayesian classifier are liking in a chain where each of them uses a Bayesian network (model of the subsection 4.3.1), for calculating the class of each affective state. As the base model has two levels, the first consisting of the MSNB of each sensor, and the second representing the class inferred by the Bayesian network, then it is possible to have chains at two-level, one at the level of MSNB (sensors level) and the other at the level of Bayesian networks (classes level). Each MSNB receives as an additional feature the classes inferred by the previous MSNB of sensors of the same type in the chain. In the same way, each Bayesian network (class) has as input variables, the classes inferred by the previous Bayesian networks in the chain. Chaining can be done at one level or at both levels at the same time.

4.4 Publishing plan

1.- Multimodal models of recognition of affective states that allow the marginalization of some sensor. Journal: *IEEE Transactions on Affective Computing*, February 2018.

2.- Multidimensional classification that includes the relations of dependence among affective states. Conference: *Affective Computing and Intelligent Interaction 2018*, deadline: Abril 30th, 2018; conference: October 2018.

3.- Multidimensional classification that includes the relations of dependence among affective states. Journal: *IEEE Transactions on Affective Computing*, February 2019.

4.5 Plan

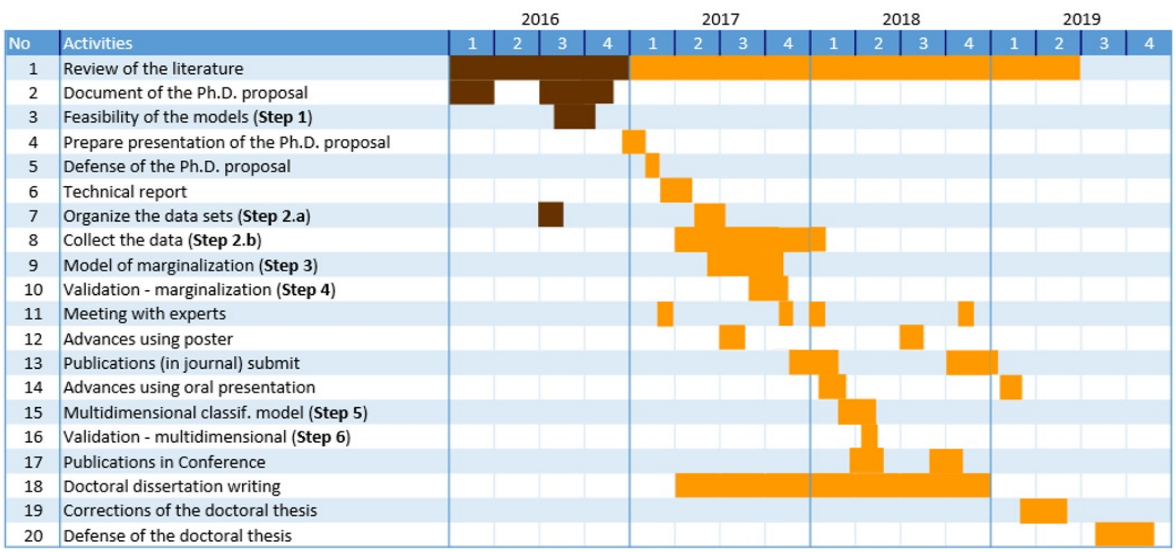


Fig. 4.7: Schedule. The specific steps described in the methodology are indicated in bold.

Chapter 5

Preliminary results

The first step in the methodology consist in developing the first versions of the models of marginalization and multidimensional classification to determine the feasibility of the proposal and compare them with the results of previous work (Rivas et al., 2016). Several experiments were carried out and the results of the three main ones are reported:

1.- Late (decision level) fusion of the sensors using Semi-Naïve Bayesian classifier.

2.- Multidimensional classification that includes dependencies among four affective states.

3.- Estimation by simple linear regression of the features of an unavailable sensor.

5.1 Dataset for preliminary experiments

The considered data set was constructed in the previous work (Rivas et al., 2016). The data set contains the records of the rehabilitation sessions of 2 stroke patients that attended therapies to recover the mobility of an upper limb. The patients (an extroverted man and an introverted woman, as judged by the psychiatrists who contributed to this study) participated in 45-minute average sessions that took place in different days (max 3 per week) in a period of 4 weeks. Demographic data and the number of rehabilitation sessions per patient are summarized in Table 5.1. The virtual rehabilitation platform *Gesture Therapy* (GT) was used in each session, and the 3D movements of the affected hand, the pressure of the fingers and a frontal video were recorded. In this process, two sensors were considered using the gripper of the GT system: one to identify the 3D location of the hand and the other to indicate the pressure exerted by the fingers (see Fig. 5.1). The frontal video was used by psychiatrists to tag the frames in which the patient exhibited one or more of the 4 considered states: tiredness (physical or psychological), anxiety, pain and engagement.

Table 5.1: Cohort demographic and data from patients' rehabilitation sessions

	Pacient $P1$	Pacient $P2$
Age [years]	55	57
Gender	M	F
Stroke date	April, 2014	May, 2014
Therapy onset	May 8th, 2014	Sep. 24th, 2014
Paretic side	right	right
No. of sessions	6	10
No. of videos	29	50

Each example of the data set has 5 features of hand movement: (averages of) speed, acceleration and differential location by the axes x, y, z, 3 features of fingers' pressure: (averages of) pressure, pressure speed and pressure acceleration, and 4 binary labels (of the set $\{-1, 1\}$), one for each affective state, indicating presence (1) or absence (-1) of the respective affective state. From the process of constructing features vectors and their

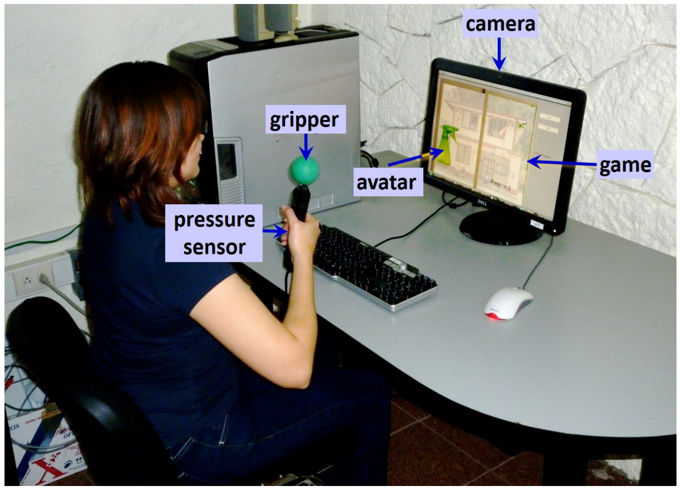


Fig. 5.1: Demonstration of the GT Platform. The gripper, held here with the right hand, is monitored with a tracking system that uses the computer camera to locate the position of the hand, and then controls an avatar in the virtual environment. The avatar is represented, in this case, by the aerosol and when the patient presses the pressure sensor with the fingers then the bottle sprays the insecticide to kill the mosquito. The mechanics of the game also induce vertical and horizontal movements of the hand (and arm) to place the bottle at the level of the mosquito. As the user interacts with the rehabilitation game, the 3D hand location, and the gripping force are recoded.

classes, a total of 5826 samples were obtained for $P1$ and 8935 for $P2$. In Fig. 5.2 the process for the construction of the samples is schematized.

Observation: For patient $P2$, the affective state of **pain was not observed** in any of her videos.

5.2 Experimental design

The base classifier in all experiments was Multiresolution Semi-Naïve Bayesian classifier (MSNB) using odd window sizes, from 3 to 11, i.e. $|V| = 3,5,7,9,11$ (see section 4.2); so that 5 different window sizes were used. The sets of examples of affective states were unbalanced with less samples of the presence of the respective affective state. Sub-sampling was applied to balance classes. Internal validation was performed using the stratified 10-fold cross-replication mechanism; randomly, the examples were distributed in such a way that each fold had the same ratio of the 1 (presence) class and the -1 (absence) class, with respect to the overall set of examples.

The classifiers MSNB and Semi-Naïve Bayesian(SNB) are binary. We calculated 6 metrics associated with the confusion matrix: $accuracy$, $sensitivity$, $specificity$, $precision$, $F - measure$, and ROC area under the curve to evaluate the performance of each classification process and compare among them.

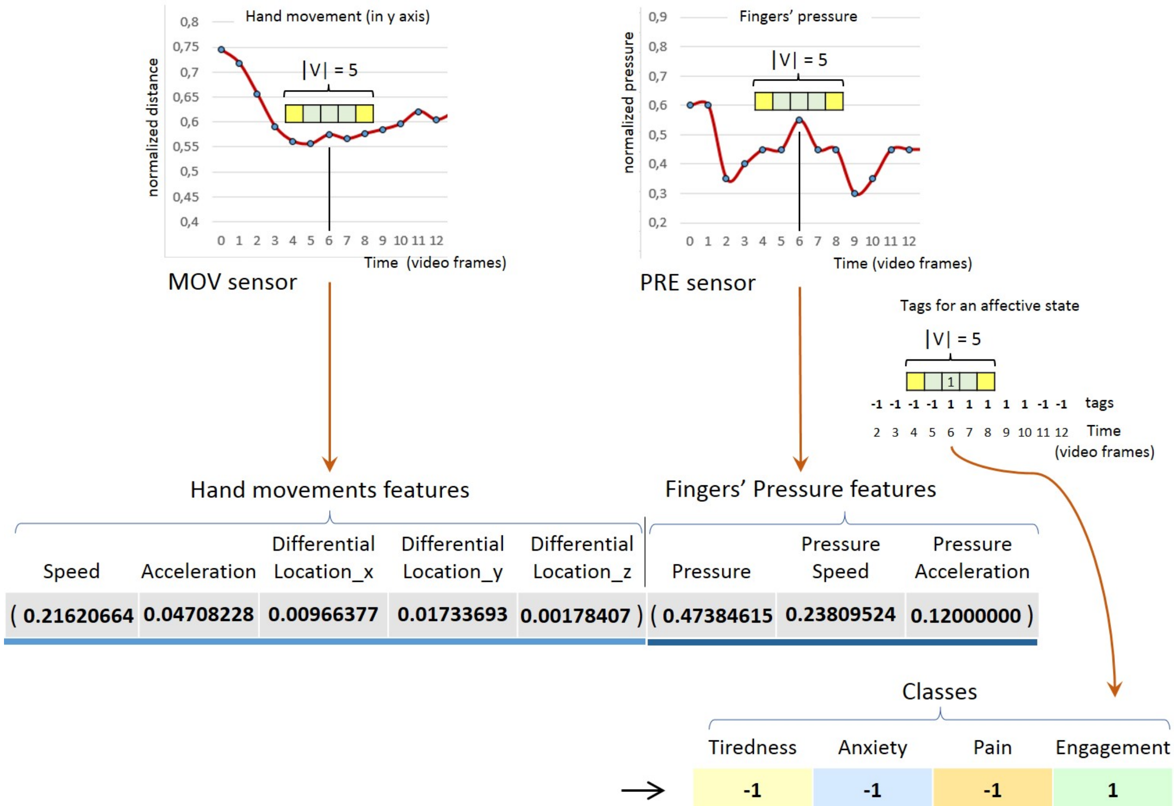


Fig. 5.2: Example of the construction of the feature vectors and their classes. In the diagram there are graphics of the traces of: hand movement (MOV sensor), finger pressure (PRE sensor) and tags of an affective state (engagement). In the traces, windows (in this case of size $|V| = 5$) are centred at the respective current point (point 6 for the example) simultaneously. The average of the feature that is being treated is calculated for the consecutive points within the window. As an example, for the speed feature of hand movement, the speed between the sub-following points within the window in hand movement trace is calculated and averaged at the end. The class (presence: 1 or absence: -1) of each affective state is assigned by determining the majority class within the window for the respective affective state trace.

5.3 Experiment 1: Late (decision level) fusion of the sensors using Semi-Naïve Bayesian classifier

A comparison of early (feature level) fusion scheme by concatenation of features, used in previous work (Rivas et al., 2016) with the late (decision level) fusion strategy proposed to be used at integration of the information from the several sensors to the models of marginalization and multidimensional classification, is made. The experimental hypothesis is: The classification performance of the late fusion model FSNB is better than the early fusion with MSNB. The objective is: Validate the preliminary model of late fusion (FSNB) with the results of MSNB of the work (Rivas et al., 2016), in which early fusion was applied (concatenation of MOV and PRE features); studying the relation among the observable variables (movements and pressure of the hand, registered with the gripper of the GT) and the four (4) states: tiredness, anxiety, pain and engagement, one at a time.

5.3.1 Design of the experiment

MSNB models with fusion at feature level (early fusion) were constructed for each affective state for each patient; for $P1$ were 4 MSNB and for $P2$ were 3 (remember $P2$ did not exhibited pain state during the labelling). In addition, FSNB models were elaborated, which have fusion at the decision level (late fusion) with the separate results of the MSNB for the MOV sensor and for the PRE sensor (see figure 5.3). During the development of the MSNB models, semi-Naïve Bayesian (SNB) models were generated, one for each window size: $|V|$ = 3,5,7,9,11.

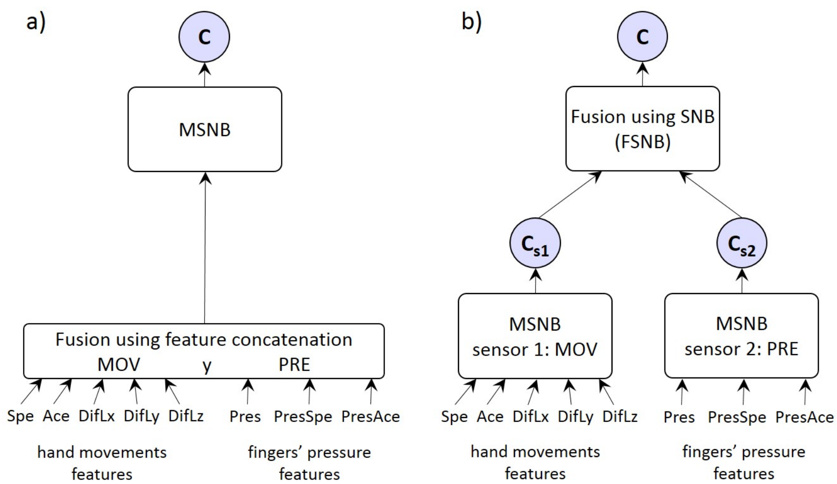


Fig. 5.3: Architectures of models of experiment 1: early fusion with MSNB and late fusion FSNB. In a) MSNB classifier with fusion at feature level is observed (the features of both sensors MOV and PRE are concatenated). In b) the FSNB model is presented, with fusion at the decision level using the SNB classifier, which receives the results of MSNB from MOV and PRE. Hand movements features are represented as: (average of) speed (Spe), acceleration (Ace), differential location x (DifLx), differential location y (DifLy) and differential location z (DifLz); fingers' features are indicated as: (average of) pressure (Pre), pressure speed (PresSpe) and pressure acceleration (PresAce).

5.3.2 Results

In Table 5.2, the average classification results are presented for both early fusion with MSNB and for late fusion with FSNB, by affective state, and the last row of each case corresponds to the average overall result. Results are summarized as mean ± standard deviation, across the 10 folds of cross-validation; for the last row through all affective states too. Pain state was the best classified for $P1$ in both MSNB and FSNB. For $P2$ the state of tiredness has the best recognition for the two strategies. The results of FSNB for $P1$ and $P2$ were significantly higher than those of MSNB using the ROC area of affective states (Wilcoxon test of sign ranges: $W = -2.618$, $p < 0.05$).

In the figure: 5.4 we can see the comparative boxplots among the average results (on affective states) of the metrics: accuracy, F-Measure and ROC area. FSNB had better

Table 5.2: Results ($\mu \pm \sigma$) of early fusion with MSNB and late fusion FSNB - Affective states, through the 10 folds of cross-validation. The highest results per column are indicated with bold typeface. The global average results for MSNB and FSNB are respectively compared and the greater one is indicated in bold

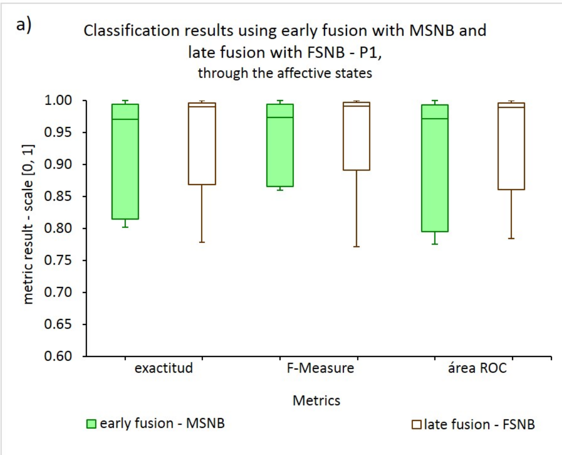
Paciente $P1$							Paciente $P2$						
Affective state	**accuracy**	**sensitivity**	**specificity**	**precision**	**F-measure**	**ROC area**	**Affective state**	**accuracy**	**sensitivity**	**specificity**	**precision**	**F-measure**	**ROC area**
MSNB							**MSNB**						
tiredness	0.926 ± 0.107	0.986 ± 0.045	0.848 ± 0.263	0.915 ± 0.127	0.943 ± 0.074	0.917 ± 0.129	tiredness	**0.929**± 0.075	**0.971** ± 0.060	0.886 ± 0.148	0.908 ± 0.109	**0.935**± 0.066	**0.929**± 0.075
anxiety	0.909 ± 0.122	**1,000** ± 0.000	0.810 ± 0.255	0.878 ± 0.146	0.928 ± 0.090	0.905 ± 0.128	anxiety	0.682 ± 0.098	0.963 ± 0.056	0.366 ± 0.235	0.640 ± 0.087	0.764 ± 0.057	0.665 ± 0.103
pain	**0.956**± 0.141	**1.000** ± 0.000	**0.943** ± 0.181	0.933 ± 0.211	0.950 ± 0.158	**0.971**± 0.090	pain	–	–	–	–	–	–
engagement	0.915 ± 0.172	0.989 ± 0.020	0.812 ± 0.384	0.915 ± 0.170	0.944 ± 0.112	0.900 ± 0.201	engagement	0.621 ± 0.093	0.824 ± 0.148	0.420 ± 0.292	0.613 ± 0.113	0.684 ± 0.057	0.622 ± 0.092
average	0.926 ± 0.133	**0.994** ± 0.025	0.853 ± 0.275	0.910 ± 0.161	0.941 ± 0.109	0.923 ± 0.140	average	0.744 ± 0.160	**0.920** ± 0.117	0.557 ± 0.326	0.720 ± 0.169	**0.794**± 0.121	0.738 ± 0.163
FSNB							**FSNB**						
tiredness	0.951 ± 0.070	0.969 ± 0.065	0.920 ± 0.140	0.951 ± 0.083	**0.957**± 0.058	0.945 ± 0.079	tiredness	0.921 ± 0.092	0.900 ± 0.151	**0.943** ± 0.074	**0.941**± 0.079	0.915 ± 0.106	0.921 ± 0.092
anxiety	0.926 ± 0.109	0.976 ± 0.039	0.873 ± 0.197	0.910 ± 0.134	0.938 ± 0.088	0.924 ± 0.112	anxiety	0.760 ± 0.081	0.875 ± 0.077	0.623 ± 0.233	0.744 ± 0.107	0.796 ± 0.054	0.749 ± 0.094
pain	**0.956**± 0.141	**1,000** ± 0.000	**0.943** ± 0.181	0.933 ± 0.211	0.950 ± 0.158	**0.971**± 0.090	pain	–	–	–	–	–	–
engagement	0.917 ± 0.163	0.927 ± 0.194	0.905 ± 0.270	**0.953**± 0.128	0.922 ± 0.157	0.916 ± 0.167	engagement	0.624 ± 0.090	0.696 ± 0.208	0.552 ± 0.279	0.645 ± 0.122	0.635 ± 0.109	0.624 ± 0.089
average	**0.937**± 0.122	0.968 ± 0.104	**0.910** ± 0.196	**0.937**± 0.142	**0.942**± 0.119	**0.939**± **0.115**	average	**0.768**± 0.150	0.824 ± 0.176	**0.706** ± 0.269	**0.776**± 0.161	0.782 ± 0.147	**0.765**± **0.152**

performance for $P1$, and was competitive with the results of early fusion with MSNB, for $P2$.

FSNB outperform the early fusion with MSNB in learning a predictive relationship among hand movements and fingers' pressure over all affective states of $P1$, with a mean value of ROC area greater than 0.939 and with a mean of ROC area of 0.765 to $P2$. Actually, values above 0.9 were reached in all affective states of $P1$ and values higher than 0.6 for $P2$.

5.4 Experiment 2: Multidimensional classification that includes the dependence among four states

This experiment aims to verify whether FSNB chain classifiers are a good alternative for the multidimensional classification that harness the relations of dependencies among affective states. The experimental hypothesis is: FSNB chain classifiers have better performance than separate FSNBs, one for each affective state. The objective of the experiment is: Validate the results of the chain in the multidimensional classification with respect to the individual results, by affective state, of FSNB for each patient; studying the relation among the observable variables (hand movements and fingers' pressure, registered with the GT gripper) and the four (4) states: tiredness, anxiety, pain and engagement.



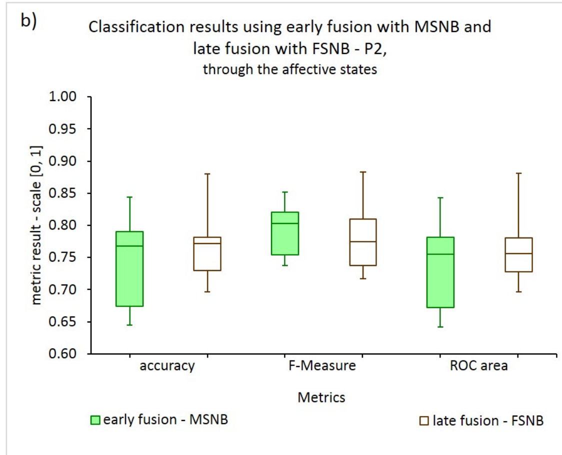


Fig. 5.4: Boxplot. Comparison of the average (on the affective states) of the results of the metrics: accuracy, F-Measure and ROC area of early fusion with MSNB and late fusion FSNB. In figure (a) it compares the boxplots for patient $P1$ and in (b) boxplots of patient $P2$.

5.4.1 Design of the experiment

In experiment 1 independent models of FSNB were developed for each of the four states (tiredness, anxiety, pain and engagement). Now the construction of multidimensional classification models is proposed through the strategy of linking FSNB in a chain. One of the first decisions is to define the affective states' order in the chain. The results of the previous work Rivas et al. (2016) allow us to locate the affective states in the sequence from the best classified to the worst:

- For $P1$: pain $\rightharpoonup$ tiredness $\rightharpoonup$ anxiety $\rightharpoonup$ engagement.
- For $P2$: tiredness $\rightharpoonup$ anxiety $\rightharpoonup$ engagement.

It was considered to construct the classification chain, for each patient, following the order described. It is assumed that the most difficult affective states to classify should be in the later posts in the chain. The classification could be favoured when these states receive information of previous affective states (see figure 5.5). In the proposed architecture, the chains can be made in two levels, sensors level and/or classes level. The chain at sensors level was only explored in this experiment (the classes level has not been studied yet), which incorporate all previous classes inferred to the corresponding sensor by base classifier MSNB.

By contrasting the results obtained for FSNB in experiment 1, it can be verified if the chain has better performance.

5.4.2 Results

In the Table 5.3 the results of FSNB chain classifier are presented and compared with the results of individual FSNB, by affective state, for $P1$ and $P2$. The last row, by classification

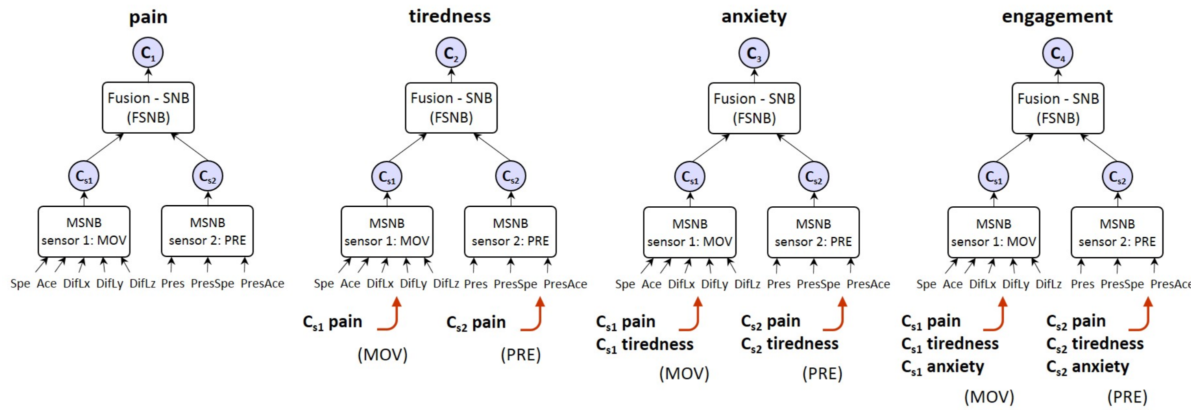


Fig. 5.5: Architecture of FSNB chain classifiers for $P1$, locating the states in the order: pain, tiredness, anxiety and engagement. The chain was studied at the sensors level, i.e. among the classes inferred to the corresponding sensor by base classifier MSNB. For example, inferred classes by MSNB from MOV sensor of previous affective states are incorporated as additional features to each MSNB that uses MOV sensor features.

model, represents the average overall performance ($\mu \pm \sigma$, across all affective states). FSNB chain classifier improved individual FSNB ROC area results for $P2$ but not for $P1$. Actually, $P1$ ROC area results are already in the order of 0.93 ± 0.115. For $P2$ there were more difficulties in individual discrimination of the affective states of anxiety and engagement; and the chain provides information that is used in these states to improve their classification. $P2$ chain results are significantly better than those of $P2$ individual FSNBs (Wilcoxon Signal Range Test: $W = -2.632$, $p < 0.05$).

The FSNB chain classifier obtained better results than individual FSNB for patient $P2$, in all metrics, but did not improve for $P1$. It is necessary to study the performance of the chain with other sequences of affective states and even to discern which states contribute more information to each other. It should also experiment with the classes level in the chain, and even with both levels at a time.

5.5 Experiment 3: Estimation by simple linear regression of the features of an unavailable sensor

One of the first explorations of the problem of prescinding from a sensor, was to eliminate it and verify if a simple linear regression strategy could be used to estimate the features of that sensor in terms of the features of the other sensor. This is the first approach to the problem of marginalization. It is clear that this approach can be further improved by using other, more sophisticated, marginalization strategies. The experimental hypothesis is: Through the feature values of the available sensor can be estimated approximations for the feature values of the missing sensor, by simple linear regression, and these estimations can be used as inputs for MSNB of the missing sensor and the FSNB model can achieve better classification than the one obtained only with the features of the available sensor.

Table 5.3: Results ($\mu \pm \sigma$) individual FSNB vs FSNB chain classifiers - Affective states, through the 10 folds of cross-validation. The highest results per column are indicated with bold typeface.

Pacient $P1$							Pacient $P2$						
Affective state	**accuracy**	**sensitivity**	**specificity**	**precision**	**F-measure**	**ROC area**	**Affective state**	**accuracy**	**sensitivity**	**specificity**	**precision**	**F-measure**	**ROC area**
FSNB							**FSNB**						
tiredness	0.951 ± 0.070	0.969 ± 0.065	0.920 ± 0.140	0.951 ± 0.083	**0.957**± 0.058	0.945 ± 0.079	tiredness	**0.921**± 0.092	**0.900** ± 0.151	**0.943** ± 0.074	**0.941**± 0.079	**0.915**± 0.106	**0.921**± 0.092
anxiety	0.926 ± 0.109	0.976 ± 0.039	0.873 ± 0.197	0.910 ± 0.134	0.938 ± 0.088	0.924 ± 0.112	anxiety	0.760 ± 0.081	0.875 ± 0.077	0.623 ± 0.233	0.744 ± 0.107	0.796 ± 0.054	0.749 ± 0.094
pain	**0.956**± 0.141	**1,000** ± 0.000	0.943 ± 0.181	0.933 ± 0.211	0.950 ± 0.158	**0.971**± 0.090	pain	–	–	–	–	–	–
engagement	0.917 ± 0.163	0.927 ± 0.194	0.905 ± 0.270	0.953 ± 0.128	0.922 ± 0.157	0.916 ± 0.167	engagement	0.624 ± 0.090	0.696 ± 0.208	0.552 ± 0.279	0.645 ± 0.122	0.635 ± 0.109	0.624 ± 0.089
average	**0.937**± 0.122	**0.968** ± 0.104	0.910 ± 0.196	**0.937**± 0.142	**0.942**± 0.119	**0.939**± 0.115	average	0.768 ± 0.150	0.824 ± 0.176	0.706 ± 0.269	0.776 ± 0.161	0.782 ± 0.147	0.765 ± 0.152
Chain							**Chain**						
tiredness	0.951 ± 0.070	0.971 ± 0.090	0.928 ± 0.135	0.951 ± 0.083	0.956 ± 0.062	0.949 ± 0.074	tiredness	**0.921**± 0.092	**0.900** ± 0.151	**0.943** ± 0.074	**0.941**± 0.079	**0.915**± 0.106	**0.921**± 0.092
anxiety	0.909 ± 0.129	**1.000** ± 0.000	0.810 ± 0.269	0.880 ± 0.150	0.929 ± 0.093	0.905 ± 0.135	anxiety	0.784 ± 0.052	0.861 ± 0.082	0.696 ± 0.147	0.770 ± 0.076	0.808 ± 0.040	0.779 ± 0.057
pain	**0.956**± 0.141	**1.000** ± 0.000	0.943 ± 0.181	0.933 ± 0.211	0.950 ± 0.158	**0.971**± 0.090	pain	–	–	–	–	–	–
engagement	0.916 ± 0.166	0.882 ± 0.231	**0.963**± 0.105	**0.963**± 0.109	0.909 ± 0.180	0.923 ± 0.154	engagement	0.677 ± 0.093	0.744 ± 0.163	0.609 ± 0.205	0.669 ± 0.105	0.691 ± 0.102	0.677 ± 0.093
average	0.933 ± 0.128	0.963 ± 0.129	**0.911** ± 0.186	0.932 ± 0.144	0.936 ± 0.128	0.937 ± 0.116	average	**0.794**± 0.129	**0.835** ± 0.148	**0.749** ± 0.205	**0.793**± 0.142	**0.804**± 0.126	**0.792**± 0.129

The objective of the experiment is: Validate the results of trained FSNB with estimations of the features of one of the sensors not available (MOV or PRE) with respect to the results of FSNB developed only with the available sensor, and the results of FSNB developed with all the sensors (MOV and PRE).

5.5.1 Design of the experiment

The problem of removing a sensor was explored by constructing FSNB models that received two types of outputs from the multiresolution classifiers (MSNB): outputs obtained for the available sensor and outputs generated over estimations for the missing sensor. The estimations were made at the level of the features of the missing sensor, using information from the features of the available sensor and following a simple linear regression strategy among each available feature and each feature to be estimated. The Pearson correlation coefficient was calculated among each feature of the available sensor and each feature of the missing sensor, choosing the relation that had the highest value. Subsequently, the corresponding linear models were created and with them the values of the features of the missing sensor were produced (see figure 5.6). The results of FSNB models obtained by elimination of one of the sensors were compared with the results of FSNB developed only with the available sensor and the results of FSNB developed with the two sensors.

5.5.2 Results

The comparison of the results of the elimination of MOV sensor for patient $P1$ with respect to the ideal case when the two sensors are available and with the base case in which only

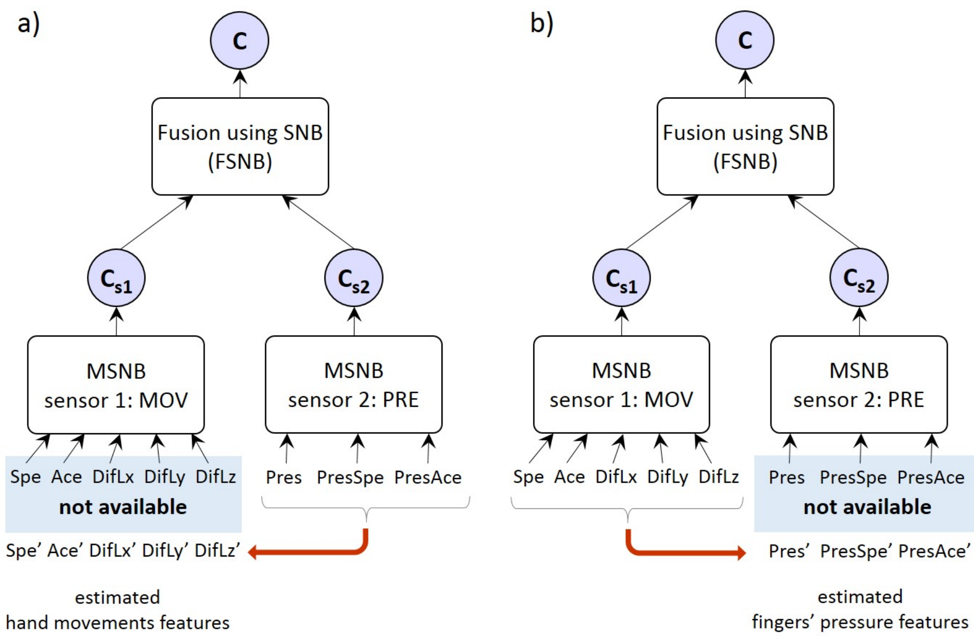


Fig. 5.6: Strategy to estimate the feature values of the sensor that is not available. In a) the MOV sensor is not available and each of its feature values were estimated through simple linear regression from the most associated feature of PRE; these features estimations of MOV and available features of PRE are used by MSNB to obtain the class C_{s1}. In b) the opposite case is shown when PRE is not available. Hand movements features are represented as: (average of) speed (Spe), acceleration (Ace), differential location x (DifLx), differential location y (DifLy) and differential location z (DifLz); fingers' features are indicated as: (average of) pressure (Pre), pressure speed (PresSpe) and pressure acceleration (PresAce).

PRE sensor is available, is developed in Table 5.4 where FSNB MOV-PRE represents the FSNB results for the two, MOV and PRE sensors; MOV-estim represents the results of FSNB with MOV estimation and FSNB PRE, the results of FSNB of PRE sensor. In all states, except engagement, and in the average over states, the ROC area results of MOV removal are between the ROC area results of the case when the two sensors are available and the ROC area results of the trivial marginalization. For the pain state the MOV removal reached the closest values to the ideal case, on the contrary for the anxiety state the results were the same as the base case: FSNB PRE. On average the simple linear regression strategy generated good results for the removal of MOV for patient $P1$.

Table 5.4: Results ($\mu \pm \sigma$) in FSNB MOV-PRE, MOV-estim y FSNB PRE - Pacient $P1$ - Affective states, through the 10 folds of cross-validation.

Affective state	Method	Accuracy	Sensitivity	Specificity	Precision	*F-Measure*	ROC area
tiredness	FSNB MOV-PRE	0.951 ± 0.070	0.969 ± 0.065	0.920 ± 0.140	0.951 ± 0.083	0.957 ± 0.058	0.945 ± 0.079
	MOV-estim	0.868 ± 0.178	0.957 ± 0.096	0.780 ± 0.346	0.878 ± 0.192	0.899 ± 0.126	0.869 ± 0.164
	FSNB PRE	0.860 ± 0.186	0.957 ± 0.096	0.760 ± 0.375	0.872 ± 0.200	0.894 ± 0.130	0.859 ± 0.178
anxiety	FSNB MOV-PRE	0.926 ± 0.109	0.976 ± 0.039	0.873 ± 0.197	0.910 ± 0.134	0.938 ± 0.088	0.924 ± 0.112
	MOV-estim	0.858 ± 0.172	1.000 ± 0.000	0.704 ± 0.360	0.828 ± 0.182	0.895 ± 0.117	0.852 ± 0.180
	FSNB PRE	0.858 ± 0.172	1.000 ± 0.000	0.704 ± 0.360	0.828 ± 0.182	0.895 ± 0.117	0.852 ± 0.180
pain	FSNB MOV-PRE	0.956 ± 0.141	1.000 ± 0.000	0.943 ± 0.181	0.933 ± 0.211	0.950 ± 0.158	0.971 ± 0.090
	MOV-estim	0.894 ± 0.193	1.000 ± 0.000	0.779 ± 0.376	0.875 ± 0.237	0.913 ± 0.178	0.889 ± 0.188
	FSNB PRE	0.872 ± 0.254	0.900 ± 0.316	0.779 ± 0.376	0.847 ± 0.319	0.869 ± 0.313	0.839 ± 0.296
engagement	FSNB MOV-PRE	0.917 ± 0.163	0.927 ± 0.194	0.905 ± 0.270	0.953 ± 0.128	0.922 ± 0.157	0.916 ± 0.167
	MOV-estim	0.909 ± 0.180	0.914 ± 0.199	0.905 ± 0.270	0.949 ± 0.139	0.916 ± 0.168	0.909 ± 0.184
	FSNB PRE	0.906 ± 0.189	0.844 ± 0.329	0.991 ± 0.012	0.994 ± 0.008	0.867 ± 0.284	0.918 ± 0.162
average	FSNB MOV-PRE	0.937 ± 0.122	0.968 ± 0.104	0.910 ± 0.196	0.937 ± 0.142	0.942 ± 0.119	0.939 ± 0.115
	MOV-estim	0.882 ± 0.175	0.968 ± 0.112	0.792 ± 0.335	0.883 ± 0.188	0.906 ± 0.144	0.880 ± 0.173
	FSNB PRE	0.874 ± 0.196	0.925 ± 0.232	0.808 ± 0.327	0.885 ± 0.211	0.904 ± 0.169	0.867 ± 0.205

In the Table 5.5 the results for the patient $P2$ are presented in an analogous way of $P1$ results. In all states and in the average over states, the ROC area results of MOV removal are between the ROC area results of the ideal case and the ROC area results of the base case. For the state of tiredness the elimination of MOV reached the values closest to the ideal case. On average the simple linear regression strategy generated good results for MOV removal for patient $P2$.

Table 5.5: Results ($\mu \pm \sigma$) in FSNB MOV-PRE, MOV-estim y FSNB PRE - Pacient $P2$ - Affective states, through the 10 folds of cross-validation.

Affective state	Method	Accuracy	Sensitivity	Specificity	Precision	*F-Measure*	ROC area
tiredness	FSNB MOV-PRE	0.921 ± 0.092	0.900 ± 0.151	0.943 ± 0.074	0.941 ± 0.079	0.915 ± 0.106	0.921 ± 0.092
	MOV-estim	0.907 ± 0.112	0.871 ± 0.157	0.943 ± 0.100	0.940 ± 0.111	0.900 ± 0.122	0.907 ± 0.112
	FSNB PRE	0.900 ± 0.108	0.871 ± 0.157	0.929 ± 0.101	0.927 ± 0.111	0.893 ± 0.117	0.900 ± 0.108
anxiety	FSNB MOV-PRE	0.760 ± 0.081	0.875 ± 0.077	0.623 ± 0.233	0.744 ± 0.107	0.796 ± 0.054	0.749 ± 0.094
	MOV-estim	0.720 ± 0.095	0.856 ± 0.082	0.563 ± 0.187	0.697 ± 0.099	0.764 ± 0.073	0.710 ± 0.101
	FSNB PRE	0.661 ± 0.089	0.896 ± 0.075	0.389 ± 0.192	0.626 ± 0.065	0.736 ± 0.063	0.642 ± 0.105
pain	FSNB MOV-PRE	–	–	–	–	–	–
	MOV-estim	–	–	–	–	–	–
	FSNB PRE	–	–	–	–	–	–
engagement	FSNB MOV-PRE	0.624 ± 0.090	0.696 ± 0.208	0.552 ± 0.279	0.645 ± 0.122	0.635 ± 0.109	0.624 ± 0.089
	MOV-estim	0.596 ± 0.151	0.558 ± 0.249	0.634 ± 0.212	0.598 ± 0.176	0.560 ± 0.198	0.596 ± 0.151
	FSNB PRE	0.570 ± 0.170	0.551 ± 0.311	0.588 ± 0.242	0.577 ± 0.202	0.529 ± 0.227	0.570 ± 0.170
average	FSNB MOV-PRE	0.768 ± 0.150	0.824 ± 0.176	0.706 ± 0.269	0.776 ± 0.161	0.782 ± 0.147	0.765 ± 0.152
	MOV-estim	0.741 ± 0.175	0.762 ± 0.225	0.713 ± 0.236	0.745 ± 0.195	0.741 ± 0.196	0.738 ± 0.177
	FSNB PRE	0.710 ± 0.187	0.773 ± 0.255	0.635 ± 0.290	0.710 ± 0.206	0.719 ± 0.211	0.704 ± 0.192

The model with estimation of the MOV sensor values (where MOV was marginalized) obtained results significantly higher than those of trivial marginalization, using the ROC area of affective states (Wilcoxon test of sign ranges: $W = -3.823$, $p < 0.05$).

When PRE sensor was removed, the results were the same as the base case: those of FSNB developed only with MOV sensor. In this case it is evident that the estimation of the feature values of PRE sensor from the features of MOV sensor was not good enough to replace the original values of PRE sensor.

The estimation of the feature values of MOV sensor, using simple linear regression from the features of PRE sensor, generated results that improved those of the base case (when only PRE sensor is available). On the other hand, the estimations of the feature values of PRE sensor did not produce results that exceeded the base case when only MOV sensor is available; This result reveals that PRE is more difficult to estimate by simple linear regression from MOV.

5.6 Final comments

Late fusion using SNB significantly improved the results of early fusion with MSNB for both patients. The second experiment showed the feasibility of improving the results of the classification of affective states, using the relations of dependence among them. The strategy of FSNB chain classifier evidenced to be beneficial. The results for patient $P2$ significantly enhanced, in the ROC area, those obtained by FSNB in each affective state independently. The results of patient $P1$ were similar to those achieved by independent FSNB classifiers for each affective state. $P1$ already had results of 0.93 ± 0.115 in the ROC area. Other experiments should be done by changing the sequence of the affective states in the chain, to determine how it affects the results. For the problem of sensor marginalization, an experiment was developed using a simple strategy based on linear regression to estimate the feature values of the missing sensor. This experiment was done with the intention of exploring the use of feature values estimation of a sensor based on the others and determine if this could be useful to give a solution to the problem of not having sensors in testing phase (learning using privileged information); of course, this strategy can be improved by considering dependencies at different levels and by using more sophisticated techniques such as Bayesian networks. MOV features, estimated from PRE, were beneficial to improve the classification compared to the case where only the features of PRE are available. In the opposite case, estimating PRE from MOV did not provide sufficient information. The marginalization problem will be addressed with the architecture of the FSNB classifier but changing, at the late (decision level) fusion, the SNB classifier by a Bayesian network, as explained in the methodology (see figure 4.4). The nature of the Bayesian networks allows marginalization, in fact, when there are not information about the values of some variables, partial abduction is applied to know the most probable value of the class given the evidence (the variables that have assigned values) and these variables that have no value are marginalized in the process.

From the related work chapter it is concluded that the works closest to our proposal (Wang et al., 2015d; Chen et al., 2016) use among their strategies the SVM classifier. Among the experiments carried out in this first stage of the research, which were not reported in this document, one was carried out where a model similar to the late fusion of experiment 1 (FSNB model) was constructed, but instead of SNB, SVM was placed as a classifier in the fusion. The obtained results were that the proposed model FSNB obtained better results than the one of SVM. Additionally, in previous work Rivas et al. (2016) SVM was compared with MSNB in the problem of recognition of affective states; and MSNB obtained significantly higher results, in the ROC area, than those of SVM.

These experiments and preliminary results represent the first efforts to show that the proposed models are promising.

Chapter 6

Conclusions and future work

6.1 Conclusions

At this stage of the research the literature has been reviewed and it has been shown that the work proposed is novel and has practical applications. The contributions consist in providing models that enhance the development of affective computing systems that include sensor marginalization, and whose classification is favoured when considering the relations of dependence of affective states. This last aspect can give greater robustness to the automatic recognition of affective states.

Preliminary results reveal that the proposed models of FSNB chain classifiers are promising. This model also has the advantage of building chains at two levels: sensors level and/or classes level. So far tests have been conducted at the sensor level using MSNB, and results are already encouraging. The problem of marginalization has not yet been fully addressed; only the problem of the absence of a sensor at testing phase has been explored. The chosen strategy, in a two sensor problem, was to estimate the values of the features of the missing sensor through the features of the available sensor. The performance of the classification was mixed, for a sensor the estimation of its features contributed to improve the classification; but the results for the other one did not leveraged the performance.

6.2 Future work

In the next step, we will study a Bayesian network as an element of fusion, at the decision level, of the proposed architecture to integrate the information of the different sensors. It is necessary to know the performance of this classifier and compare with the results obtained so far with FSNB. The architecture including the Bayesian network, as integrator of sensors, would favour to tackle the problem of marginalization of some sensor.

The proposed architecture of FSNB chain classifiers, to consider the dependence relations of affective states, has been only studied at the level of the sensors; it is necessary to experiment with the chain at classes level, and then include the two chain levels (sensors and classes) at the same time.

The other aspect to be addressed is to continue the organization of the datasets (step 2 of the methodology). DEAP and MAHNOB-HCI datasets have already been acquired; but it is required to begin the process for the construction of the own data set.

Bibliography

Bailenson, J. N., Pontikakis, E. D., Mauss, I. B., Gross, J. J., Jabon, M. E., Hutcherson, C. A., Nass, C., and John, O. (2008). Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International journal of human-computer studies*, 66(5):303–317.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., and Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, 78(5):B231–B244.

Betta, J. C. (1974). *Manual de psiquiatría*. Editorial Albatros.

Bhatia, M. S. (2009). *Dictionary of psychology and allied sciences*. New Age International.

Bianchi-Berthouze, N., Kim, W., and Patel, D. (2007). Does body movement engage you more in digital game play? and Why? *Affective computing and intelligent interaction*, pages 102–113.

Bielza, C., Li, G., and Larranaga, P. (2011). Multi-dimensional classification with Bayesian networks. *International Journal of Approximate Reasoning*, 52(6):705–727.

Bonarini, A., Garbarino, M., Matteucci, M., and Tognetti, S. (2010). Affective evaluation of robotic rehabilitation of upper limbs in post-stroke subjects. In *Proceedings 1st Workshop on the Life Sciences at Politecnico di Milano*, pages 290–293.

Cabanac, M. (2002). What is emotion? *Behavioural processes*, 60(2):69–83.

Calvo, R. A., Brown, I., and Scheding, S. (2009). Effect of experimental factors on the recognition of affective mental states through physiological measures. In *Australasian Joint Conference on Artificial Intelligence*, pages 62–70. Springer.

Carrizosa, E. and Martín-Barragán, B. (2011). Problemas de clasificación: problemas de localización. Technical report, Facultad de Matemáticas, Universidad de Sevilla.

Carrizosa, E., Martín-Barragán, B., and Morales, D. R. (2011). Detecting relevant variables and interactions in supervised classification. *European Journal of Operational Research*, 213(1):260–269.

Cernea, D. and Kerren, A. (2015). A survey of technologies on the rise for emotion-enhanced interaction. *Journal of Visual Languages & Computing*, 31:70–86.

Chanel, G., Rebetez, C., Bétrancourt, M., and Pun, T. (2011). Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41(6):1052–1063.

Chen, L. S., Huang, T. S., Miyasato, T., and Nakatsu, R. (1998). Multimodal human emotion/expression recognition. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 366–371. IEEE.

Chen, S., Wang, S., Wu, C., Gao, Z., Shi, X., and Ji, Q. (2016). Implicit hybrid video emotion tagging by integrating video content and users' multiple physiological responses. In *Proceedings of the 23rd International Conference on Pattern Recognition (ICPR)*, pages 290–295. IEEE.

Chow, C. and Liu, C. (1968). Approximating discrete probability distributions with dependence trees. *Information Theory, IEEE Transactions on*, 14(3):462–467.

Chumbley, J. and Griffiths, M. (2006). Affect and the computer game player: the effect of gender, personality, and game reinforcement structure on affective responses to computer game-play. *CyberPsychology & Behavior*, 9(3):308–316.

Clynes, M. and Menuhin, Y. (1977). *Sentics: The touch of emotions.* Anchor Press Garden City, NY.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, 16(7-8):555–575.

Consuegra, N. (2010). *Diccionario de psicología.* 2da edition.

Costa, P. T. and MacCrae, R. R. (1992). *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO FFI): Professional manual.* Psychological Assessment Resources.

Ekman, P., Levenson, R. W., and Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221(4616):1208–1210.

Ffrench, G. (1960). The clinical significance of tiredness. *Canadian Medical Association journal*, 82(13):665.

Gao, Y., Bianchi-Berthouze, N., and Meng, H. (2012). What does touch tell us about emotions in touchscreen-based gameplay? *ACM Transactions on Computer-Human Interaction (TOCHI)*, 19(4):31.

Gao, Z. and Wang, S. (2015). Emotion recognition from EEG signals using hierarchical Bayesian network with privileged information. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pages 579–582. ACM.

Gilleade, K. and Allanson, J. (2003). A toolkit for exploring affective interface adaptation in videogames.

Gilleade, K. M. and Dix, A. (2004). Using frustration in the design of adaptive videogames. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*, pages 228–232. ACM.

Godbole, S. and Sarawagi, S. (2004). Discriminative methods for multi-labeled classification. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 22–30. Springer.

Goldberg, B., Sottilare, R., Brawner, K., and Holden, H. (2011). Predicting learner engagement during well-defined and ill-defined computer-based intercultural interactions. *Affective Computing and Intelligent Interaction*, pages 538–547.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. *Review of personality and social psychology*, 2(1):141–165.

Goldberg, L. R. (1990). An alternative "description of personality": the big-five factor structure. *Journal of personality and social psychology*, 59(6):1216–1229.

Gunes, H., Piccardi, M., and Pantic, M. (2008). *From the lab to the real world: Affect recognition using multiple cues and modalities.* InTech Education and Publishing.

Hart, P. M., Caballero, C. L., and Cooper, W. (2010). Understanding engagement: its structure, antecedents and consequences. In *International Academy of Management and Business Summer Conference, Madrid.*

Hu, J. (2008). *Data Fusion: A First Step in Decision Informatics.* ProQuest.

Jeswani, S. and Dave, S. (2012). Impact of individual personality on turnover intention a study on faculty members. *Management and Labour Studies*, 37(3):253–265.

Kim, J., André, E., and Vogt, T. (2009). Towards user-independent classification of multimodal emotional signals. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–7. IEEE.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., and Patras, I. (2012). DEAP: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31.

Little, R. J. and Rubin, D. B. (2014). *Statistical analysis with missing data.* John Wiley & Sons.

Luneski, A., Bamidis, P. D., and Hitoglou-Antoniadou, M. (2008). Affective computing and medical informatics: state of the art in emotion-aware medical applications. *Studies in Health Technology and Informatics*, 136:517–522.

Luneski, A., Konstantinidis, E., and Bamidis, P. (2010). Affective medicine: a review of affective computing efforts in medical informatics. *Methods of information in medicine*, 49(3):207–218.

Mangai, U. G., Samanta, S., Das, S., and Chowdhury, P. R. (2010). A survey of decision fusion and feature fusion strategies for pattern classification. *IETE Technical review*, 27(4):293–307.

Martínez Arrollo, M. (2007). *Aprendizaje de clasificadores bayesianos estáticos y dinámicos.* PhD thesis, Instituto Tecnológico y de Estudios Superiores de Monterrey.

Martínez-Arroyo, M. and Sucar, L. E. (2006). Learning an optimal naive Bayes classifier. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 3, pages 1236–1239. IEEE.

Matsumoto, D. (2009). *The Cambridge dictionary of psychology.* Cambridge University Press.

Mehrabian, A. (1980). Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies. *Oelgeschlager, Gunn & Hain, Cambridge, MA*, pages 39–53.

Mitchell, T. M. (1997). *Machine learning. WCB.* McGraw-Hill Boston, MA:.

Molins-Ruano, P., Sevilla, C., Santini, S., Haya, P., Rodríguez, P., and Sacha, G. (2014). Designing videogames to improve students' motivation. *Computers in Human Behavior*, 31:571–579.

Muñoz Rosas, J. F. and Alvarez Verdejo, E. (2009). Métodos de imputación para el tratamiento de datos faltantes: aplicación mediante R/Splus. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 7.

Pantic, M., Pentland, A., Nijholt, A., and Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In *Artifical Intelligence for Human Computing*, pages 47–71. Springer.

Pazzani, M. J. (1996). Searching for dependencies in Bayesian classifiers. In *Learning from Data*, pages 239–248. Springer.

Pervin, L. A. (2015). *Goal concepts in personality and social psychology.* Psychology Press.

Picard, R. W. (2003). Affective computing: challenges. *International Journal of Human-Computer Studies*, 59(1):55–64.

Picard, R. W., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(10):1175–1191.

Posner, J., Russell, J. A., and Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(03):715–734.

Read, J., Pfahringer, B., Holmes, G., and Frank, E. (2009). Classifier chains for multi-label classification. *Machine Learning and Knowledge Discovery in Databases*, pages 254–269.

Renau, V., Oberst, U., Gosling, S. D., Rusiñol, J., and Lusar, A. C. (2013). Translation and validation of the Ten-Item-Personality Inventory into Spanish and Catalan. *Aloma: revista de psicologia, ciències de l'educació i de l'esport Blanquerna*, (31 (2)):85–97.

Rivas, J. J. (2015). Clasificador Semi-Naïve Bayes con multiresolución para la estimación de estados afectivos: Aplicación en rehabilitación virtual. Master's thesis, Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE).

Rivas, J. J., Orihuela-Espina, F., Palafox, L., Bianchi-Berthouze, N., Lara, M. d. C., Hernández-Franco, J., and Sucar, L. E. (2016). Unobtrusive inference of affective states in virtual rehabilitation from upper limb motions: A feasibility study.

Romero, E., Villar, P., Gómez-Fraguela, J. A., and López-Romero, L. (2012). Measuring personality traits with ultra-short scales: A study of the Ten Item Personality Inventory (TIPI) in a Spanish sample. *Personality and Individual Differences*, 53(3):289–293.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161 – 1178.

Savva, N. and Bianchi-Berthouze, N. (2011). Automatic recognition of affective body movement in a video game scenario. In *International Conference on Intelligent Technologies for interactive entertainment*, pages 149–159. Springer.

Schafer, J. L. and Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 7(2):147.

Schaufeli, W. B., Martinez, I. M., Pinto, A. M., Salanova, M., and Bakker, A. B. (2002). Burnout and engagement in university students a cross-national study. *Journal of cross-cultural psychology*, 33(5):464–481.

Scherer, K. R. (2005). What are emotions? and how can they be measured? *Social science information*, 44(4):695–729.

Shaker, N., Yannakakis, G. N., and Togelius, J. (2010). Towards automatic personalized content generation for platform games. In *AIIDE*.

Soleymani, M., Lichtenauer, J., Pun, T., and Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1):42–55.

Soleymani, M. and Pantic, M. (2012). Human-centered implicit tagging: Overview and perspectives. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3304–3309. IEEE.

Sorower, M. S. (2010). A literature survey on algorithms for multi-label learning. *Oregon State University, Corvallis*.

Steinberg, L. (1996). Beyond the classroom: Why school reform has failed and what parents need to do. *Touchstone*.

Sucar, L. E. (2015). *Probabilistic Graphical Models: Principles and Applications*. Springer.

Sucar, L. E., Bielza, C., Morales, E. F., Hernandez-Leal, P., Zaragoza, J. H., and Larrañaga, P. (2014). Multi-label classification with Bayesian network-based chain classifiers. *Pattern Recognition Letters*, 41:14–22.

Tupes, E. C. and Christal, R. E. (1961). Recurrent personality factors based on trait ratings. Technical report, DTIC Document.

Van Reekum, C., Johnstone, T., Banse, R., Etter, A., Wehrle, T., and Scherer, K. (2004). Psychophysiological responses to appraisal dimensions in a computer game. *Cognition and emotion*, 18(5):663–688.

Vapnik, V. and Vashist, A. (2009). A new learning paradigm: Learning using privileged information. *Neural Networks*, 22(5):544–557.

Verduyn, P., Delaveau, P., Rotgé, J.-Y., Fossati, P., and Van Mechelen, I. (2015). Determinants of emotion duration and underlying psychological and neural mechanisms. *Emotion Review*, 7(4):330–335.

Vinola, C. and Vimaladevi, K. (2015). A survey on human emotion recognition approaches, databases and applications. *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, 14(2):24–44.

Wang, S., He, M., Zhu, Y., He, S., Liu, Y., and Ji, Q. (2015a). Learning with privileged information using Bayesian networks. *Frontiers of Computer Science*, 9(2):185–199.

Wang, S., Wang, J., Wang, Z., and Ji, Q. (2015b). Multiple emotion tagging for multimedia data by exploiting high-order dependencies among emotions. *IEEE Transactions on Multimedia*, 17(12):2185–2197.

Wang, S., Wang, Z., and Ji, Q. (2015c). Multiple emotional tagging of multimedia data by exploiting dependencies among emotions. *Multimedia Tools and Applications*, 74(6):1863–1883.

Wang, S., Zhu, Y., Wu, G., and Ji, Q. (2014a). Hybrid video emotional tagging using users' EEG and video content. *Multimedia tools and applications*, 72(2):1257–1283.

Wang, S., Zhu, Y., Yue, L., and Ji, Q. (2015d). Emotion recognition with the help of privileged information. *IEEE Transactions on Autonomous Mental Development*, 7(3):189–200.

Wang, Z. and Ji, Q. (2015). Classifier learning with hidden information. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4969–4977.

Wang, Z., Wang, S., He, M., Liu, Z., and Ji, Q. (2013). Emotional tagging of videos by exploring multiple emotions' coexistence. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–6. IEEE.

Wang, Z., Wang, X., and Ji, Q. (2014b). Learning with hidden information. In *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR)*, pages 238–243. IEEE.

Wrosch, C. and Scheier, M. F. (2003). Personality and quality of life: The importance of optimism and goal adjustment. *Quality of life Research*, 12(1):59–72.

Wu, S., Wang, S., Zhu, Y., Gao, Z., Yue, L., and Ji, Q. (2016). Employing subjects' information as privileged information for emotion recognition from EEG signals. In *Proceedings of the 23rd International Conference on Pattern Recognition (ICPR)*, pages 296–301. IEEE.

Yang, Y. and Webb, G. I. (2001). Proportional k-interval discretization for naive-Bayes classifiers. In *12th European Conference on Machine Learning (ECML)*, pages 564–575. Springer.

Zeng, Z., Pantic, M., and Huang, T. S. (2009). Emotion recognition based on multimodal information. In *Affective information processing*, pages 241–265. Springer.

Zhalehpour, S., Akhtar, Z., and Erdem, C. E. (2014). Multimodal emotion recognition with automatic peak frame selection. In *Innovations in Intelligent Systems and Applications (INISTA) Proceedings, 2014 IEEE International Symposium on*, pages 116–121. IEEE.

Zhao, H. and Seibert, S. E. (2006). The big five personality dimensions and entrepreneurial status: a meta-analytical review. *Journal of applied psychology*, 91(2):259.

Zheng, W.-L., Dong, B.-N., and Lu, B.-L. (2014). Multimodal emotion recognition using EEG and eye tracking data. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5040–5043. IEEE.