Dealing with a Missing Sensor in a Multilabel and Multimodal Automatic Affective States Recognition System

Jesús Joel Rivas
Computer Science Department
Instituto Nacional de Astrofísica,
Óptica y Electrónica (INAOE)
Puebla, México
jrivas@inaoep.mx

Felipe Orihuela-Espina
Computer Science Department
Instituto Nacional de Astrofísica,
Óptica y Electrónica (INAOE)
Puebla, México
f.orihuela-espina@inaoep.mx

Luis Enrique Sucar
Computer Science Department
Instituto Nacional de Astrofísica,
Óptica y Electrónica (INAOE)
Puebla, México
esucar@inaoep.mx

Nadia Bianchi-Berthouze
UCL Interaction Centre
University College of London (UCL)
London, United Kingdom
n.berthouze@ucl.ac.uk

***Abstract*—Data from multiple sensors can boost the automatic recognition of multiple affective states in a multilabel and multimodal recognition system. At any time, the streaming from any of the contributing sensors can be missing. This work proposes a method for dealing with a missing sensor in a multilabel and multimodal automatic affective states recognition system. The proposed method, called Hot Deck using Conditional Probability Tables (HD-CPT), is incorporated into a multimodal affective state recognition system for compensating the loss of a sensor using the recorded historical information of the sensor and its interaction with the other available sensors. In this work, we consider a multilabel classifier, named Circular Classifier Chain, for the automatic recognition of four states: tiredness, anxiety, pain, and engagement; combined with a multimodal classifier based on three sensors: fingers pressure, hand movements, and facial expressions; which was adapted for coping with the problem of a missing sensor in a virtual rehabilitation platform for post-stroke patients. A dataset of five post-stroke patients who attended ten longitudinal rehabilitation sessions was used for the evaluation. The inclusion of HD-CPT compensated for the loss of one sensor with results above those obtained with only the remaining sensors available. HD-CPT prevents the system from collapsing when a sensor fails, providing continuity of operation with results that attenuate the loss of the sensor. The proposed method HD-CPT can provide robustness for the naturalistic everyday use of an affective states recognition system.**

***Index Terms*—automatic affective states recognition, missing sensor, multilabel classification, classifier chains, multimodal classification, fingers pressure, hand movements, facial expressions, virtual rehabilitation**

This project was supported by a grant from the project FORDECyT 296737 "Consorcio en Inteligencia Artificial". We would also like to acknowledge support for the Scholarship No. 434867 from the Mexican Research Council CONACYT.

I. Introduction

To increase the reliability of automatic affective states recognition, affective computing systems often use multiple types of sensors for providing complementary information about a person's affective states or for solving ambiguities in the interpretation of the data. In single sensor systems, the failure of the only input stream may be catastrophic. In multi-sensor systems, when a subset of the sensors fails to operate or a certain type of sensors is not available due to contextual reasons (e.g. for privacy reasons, a patient may prefer to switch off the video camera), it is sometimes possible to operate with the remaining sensors. Consequently, it is convenient to develop solutions to allow multi-sensor (multimodal) systems to continue operating, but also important to do so to keep a reliable performance even when a sensor is missing, i.e. not merely surviving crashing. Moreover, a successful strategy should aim to compensate and keep the performance higher than the performance that would be obtained by only using the remaining available sensors while ignoring the information provided by the missing sensor. Integrating such considerations in the multimodal affective computing systems would facilitate their deployment in environments outside the laboratory (naturalistic settings).

The work presented in this paper aims to address the need to deal with a missing sensor in the context of a multilabel and multimodal automatic affective states recognition system used in a real hospital setting. The Hot Deck imputation method [1], [2] using Conditional Probability Tables (CPTs) was incorporated into a multimodal affective states recognition system for compensating the loss of a sensor using the recorded historical information of the sensor and its interaction with the other available sensors. The recognition system being considered is based on a multilabel classifier named Circular Classifier Chains (CCC) [3]–[5] used for the automatic recognition of four states: tiredness, anxiety, pain, and engagement [5], combined with a multimodal classifier called Fusion using Semi-Naïve Bayesian Classifier (FSNBC). The FSNBC is based on three sensors [6]: PRE sensor for fingers pressure, MOV sensor for hand movements, and FAE sensor for facial expressions. This system is used in a virtual rehabilitation platform for post-stroke patients. We explore how we can preserve the reliability of the system when a sensor is missing. The main novelty of the method is the use of the CPTs and its integration into a multilabel and multimodal recognition system. Although in this work we incorporated it into a particular architecture, it can be used with any multimodal system based on a probabilistic approach.

We evaluated the proposed approach by integrating it in an application for post-stroke rehabilitation [5], [7]. The dataset was recorded during ten longitudinal rehabilitation sessions using a virtual rehabilitation platform named Gesture Therapy [8], [9]. This application was chosen because it was used in a real-hospital scenario, is multimodal (facial expressions, finger pressures, and hand movements), and includes four (affective, physical, or psychological) states of the patients: tiredness, anxiety, pain, and engagement, in a multilabel scheme. By being multilabel, we could investigate the effect of the proposed approach to cope with different states, where sensors may differ in their relevance to the estimation of the different states. The four states were related to patients' physical activity sessions and were chosen through discussions with clinician staff [5], [7].

Experiments in three simulations were carried out using the dataset of post-stroke patients mentioned above. In each simulation, a different sensor was treated as lost and the performance of the system that incorporated the proposed method Hot Deck using Conditional Probability Tables (HD-CPT) was compared against the system with all the sensors available and the system that just eliminated the corresponding sensor. Our approach gave robustness to the system for continuing working although a sensor was lost, it did not represent extra computational cost and provided recognition performances that were above the results of the computational model when only the working sensors were considered.

In summary, the main contributions are the following:

1) We propose a novel extension of Hot Deck method to work with Conditional Probability Tables (HD-CPT) which can provide robustness for the naturalistic everyday use of an affective states recognition system.
2) The integration of the HD-CPT method in a multilabel and multimodal affective states recognition system.
3) The method does not add extra computational cost to the model when a sensor is lost, and in general provides better results that just ignoring the missing sensor.
4) The method was evaluated in a real world application with competitive results.

II. Related Work

The problem of learning using privileged information paradigm [10] is closely related to the problem of a missing sensor in some computational models. 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 testing or deployment phase. While this paradigm has been explored mainly in the context of Support Vector Machines (SVM), new research has emerged concerning other machine learning algorithms and in the context of affective computing [11], [12].

One of the applications has been the implicit tagging of emotional videos, in which the observer's physiological responses and nonverbal spontaneous behavior displayed when interacting with the videos are used to label segments of a video with the various emotions the video segments induce to the observer [13]. An approach of implicit video emotion tagging and recognition of affective states from Electroencephalogram (EEG) signals was developed [14] through the use of Canonical Correlation Analysis (CCA). Two new feature spaces were created, one for the EEG and one for the video, which encapsulated the relationships 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 for implicit video tagging. The SVM from the EEG feature space uses the features of the video as privileged information to recognize the person's emotional state. The experiments showed higher performances in valence and arousal classifications than simply using classical SVM classifications. A drawback of this approach is that it does not provide a solution for the case of more than two sensors.

To overcome the two-sensor limitation, a SVM model with similarity restrictions in the mapping functions [15] was developed to capture the relationship between EEG signals, multiple user peripheral physiological signals (Electrooculogram (EOG), Electromyogram (EMG), Electrocardiogram (ECG), Galvanic Skin Response (GSR), Respiration (RESP), Skin Temperature (TEMP), and Plethysmogram (PLET)) and the features of videos' 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, only the video features are available. To obtain the classification model, an optimization problem with the SVM with the similarity restrictions in the mapping functions was solved using the Lagrange multiplier and solving its dual problem, which implies high computational costs.

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 [16]. This 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 to predict binary valence and arousal levels; 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 of the EEG signals of the group of individuals in automatic recognition of emotions [17]. This approach also implies a high computational cost for generating the structure of the Bayesian networks and for obtaining their parameters.

Bayesian networks inherently allow the marginalization of features and, for this reason, are useful for learning using privileged information [18]. Bayesian structures have been studied [18] with 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. However, this produces a complex structure and implies that the inference process requires high computational resources.

The proposed method based on HD-CPT can be used in computational models with more than two sensors and with no extra computational costs. The proposed method does not represent an overhead for the multilabel and multimodal computational model (as shown in Sec. VIII). The creation of the Conditional Probability Tables (CPTs) is simple, and the process for choosing the required value is straightforward based on the Conditional Probability Table (CPT).

III. Proposed Method

The proposed method estimates the values of the missing sensor by applying the Hot Deck scheme [1] combined with a strategy of selecting the required values according to the probability distribution registered in the corresponding CPT.

A. Hot Deck

Hot Deck is a classical imputation method that replaces missing values with observed values from a "similar" feature vector [1]. The process involves replacing missing values of one or more components of a feature vector $\vec{x}_v$ (called the receiver) with observed values provided by a feature vector $\vec{x}_u$ (called the donor) which is similar to the receiver for the non-empty components [2]. The final donor can be selected randomly from a set of feature vectors that are potential donors, or it can be selected deterministically using, for example, the "nearest neighbour" [2].

B. Hot Deck using Conditional Probability Tables

Our novel methodological contribution is to extend Hot Deck to work over Conditional Probability Tables (HD-CPT). Since there should be historical information (or training data) of the values of all variables, i.e., feature vectors with values for all its components, then when a new feature vector has missing values in some of its components, it should be possible to make the imputation through the Hot Deck method using the corresponding CPT.

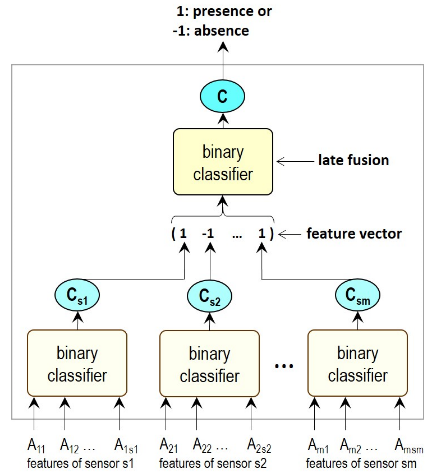

Fig. 1. Schematic depiction of a multimodal binary classifier. The features of m sensors are the inputs for m binary classifiers, respectively. Each binary classifier predicts the presence (1) or the absence (-1) of a state yielding the outputs C_{sh} = class variable for the inference from sensor sh, $h \in \{1, 2, \ldots, m\}$. The predicted classes are the components of a feature vector that is the input to a final binary classifier to performs a late fusion to predict the final binary response.

In the context of multimodal binary classification, where several sensors are involved, and the information of each sensor is processed by a corresponding binary classifier, a feature vector can be created whose components are the classes predicted by the binary classifiers. Each binary classifier predicts the presence or the absence of a state from information received from a different sensor, respectively. The feature vector is used as input to a final classifier that performs a late fusion to predict the final binary response (Fig. 1). If a sensor fails, it leads to the respective component in the feature vector having a missing value. In this context, identical donor feature vectors in the historical records can provide different values for the missing value in the receiver feature vector. In this situation, the missing value is imputed by selecting the required value according to the probability distribution registered in the corresponding CPT, given the values of the other variables. An example in a three sensors problem is depicted in Fig. 2.

IV. Integration of the Proposed Method into the Bayesian Classifiers Architecture

We tested the approach in a real case scenario of a post-stroke rehabilitation system that is based on a Bayesian classifiers architecture. This is an interesting testbed based on a probabilistic approach. We briefly present the architecture here (for details see [5], [7]).

A. Circular Classifier Chains (CCC) with Fusion using Semi-Naïve Bayesian Classifier (FSNBC)

All the computational models are binary classifiers assembled using the Semi-Naive Bayesian classifier (SNBC) [19], [20] as the core model. The data of each sensor are processed individually, providing each of them as input to

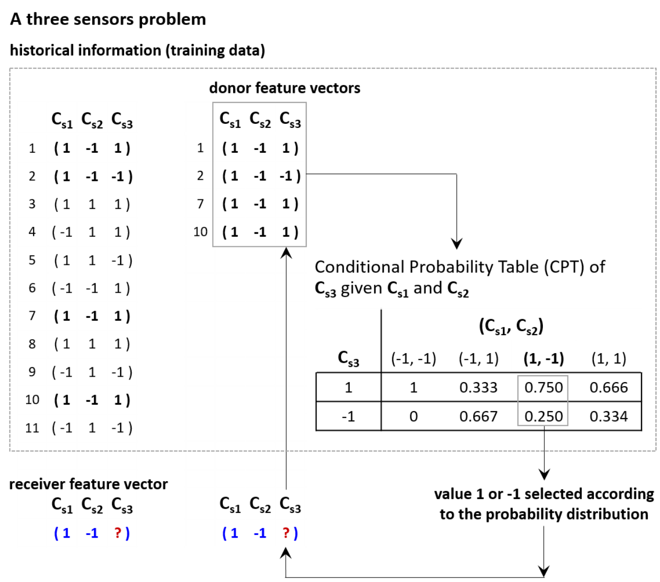

Fig. 2. An example of the proposed method: Hot Deck using Conditional Probability Tables (HD-CPT). In this example there are three sensors: s1, s2, and s3, resulting in three predicted classes: C_{s1}, C_{s2}, and C_{s3}. The historical information has 11 feature vectors. A new feature vector is presented where the third component has a missing value. The CPT of C_{s3} given C_{s1} and C_{s2} is built. For $C_{s1} = 1$ and $C_{s2} = -1$, there are 4 instances and they are the donor feature vectors. The value for C_{s3} is selected according to the probability distribution when $C_{s1} = 1$ and $C_{s2} = -1$.

a different Multiresolution Semi-Naïve Bayesian Classifier 2 (MSNBC2) [5], [7], [21] to estimate, in each case, the presence or the absence of a same affective state. Then FSNBC (late Fusion using SNBC) [6] is used for processing the prediction (occurrence or not of the affective state) from the MSNBC2 of each sensor and finally decides the occurrence of the affective state (Fig. 3).

There are as many FSNBC as affective states, each one for recognizing one affective state. These FSNBC are linked using a CCC [3]–[5]. CCC integrates the dependency relationships between the affective states to get the final recognition. CCC is a multilabel model that consists of q base binary classifiers linked in a circular chain that executes an iterative process for propagating the predicted classes to the succeeding classifiers until a fixed number of iterations is reached or until convergence. The architecture is illustrated in Fig. 3. See [3]–[5] for more details.

B. Incorporating the proposed method: HD-CPT

The proposed method provides a mechanism to deal with a missing sensor in a probabilistic classifier. We show here how it can be embedded in CCC-FSNBC. The mechanism was operationalized in the MSNBC2 component (which is part of FSNBC) since MSNBC2 processes each sensor separately. When data from a sensor are lost, the data of the other working sensors can help fill in the information of the missing sensor to minimize the reduction in recognition performance (Fig. 3). It is assumed that there is historical information of all the working sensors in the classification process, so the predicted classes of the missing sensor can be estimated with respect to the classes predicted by the other working sensors in the historical data; so it is possible to impute, through the proposed method –HD-CPT–, the classes of the missing sensor from the predicted classes of the other working sensors of that affective state.

The imputation process in the MSNBC2 of the missing sensor (the receiver) has to be made for estimating the predicted class in each odd-size sliding window W of the input signal corresponding to each sensor[1], $|W| = 3, 5, 7, 9, 11$ (the predicted class of the Semi-Naïve Bayesian Classifier (SNBC) that corresponds to each W) (Fig. 4). Then, the fusion using SNBC within MSNBC2 predicts the class that MSNBC2 provides as input to FSNBC (as the contribution from the missing sensor, which is one input modality). Specifically, when a sensor is lost, the predicted classes of the working sensors for the same odd-size sliding window $|W|$, in the corresponding affective state, can provide the information that HD-CPT needs to identify the donor vectors in the historical records. Once the donor vectors are identified, then the imputation for the predicted class in the respective odd-size window $|W|$ of the receiver can be done by selecting the class, 1 or -1, according to the probability distribution of the receiver classes given the classes of the donor vectors. Indeed, the CPT of the classes of the receiver given the classes of the donors is generated, and the selection of the class for the receiver is made according to the probability distribution registered in the CPT. The imputation process must be done for the class of each odd-size window of the receiver, in this case, five times.

After the imputation process has been made at the MSNBC2s of the missing sensor, then the architecture of CCC-FSNBC may attenuate the loss of the sensor at the subsequence levels of FSNBC and CCC when the dependency relationships of affective states are added in.

V. Virtual Neurorehabilitation

Neurorehabilitation is the clinical process that patients with sequelae of neural injury follow, aiming to recover or compensate their former function. Virtual rehabilitation is a submodality of occupational therapy where the neurorehabilitation exercises occur with the patient immersed in a virtual environment providing safety, opportune feedback, and motivation [9]. There already exist several virtual rehabilitation environments [9] with different virtues and limitations.

Gesture Therapy (GT) [5], [8], [9] is a virtual rehabilitation platform designed to attend post-stroke patients in the recovery of the mobility of their upper limbs. Therapeutic exercises are disguised in the dynamics of reaching targets in serious games that promote the mobility of the arm and hand, and finger pressure. GT is controlled by means of a gripper and a system for tracking the gripper's colour ball and for registering the finger pressure exerted on the pressure sensor in the gripper

[1]Each sensor provides a continuous input signal which is sampled, and the samples are processed in windows of different sizes for classification [23].

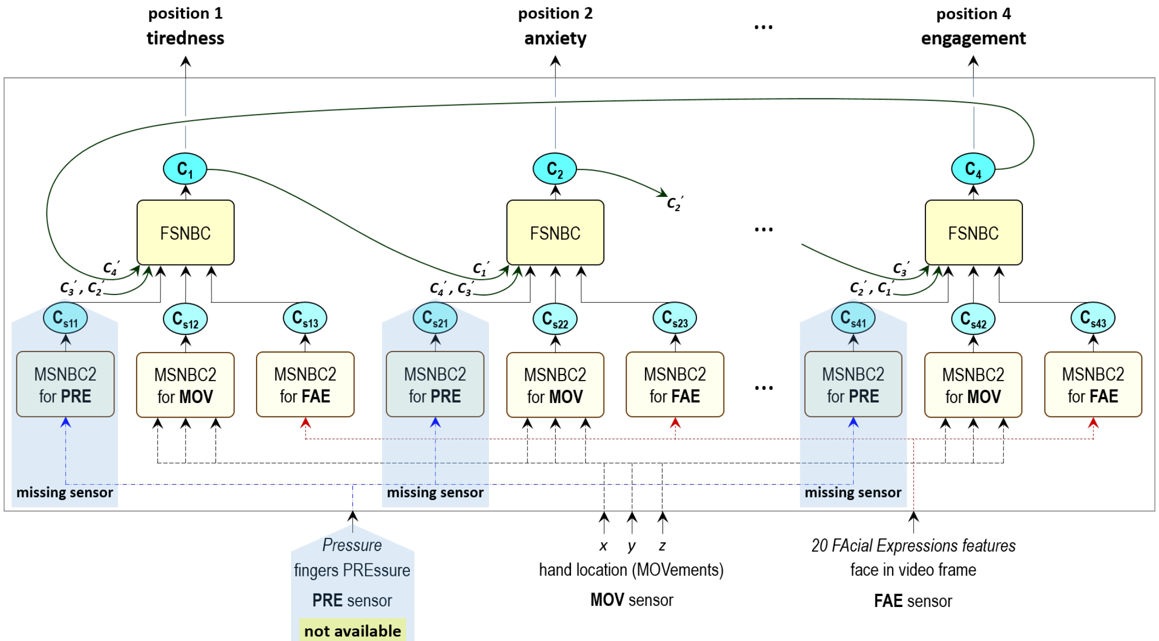

Fig. 3. Circular Classifier Chains (CCC) - Fusion using Semi-Naïve Bayesian Classifier (FSNBC), with the method of Hot Deck using Conditional Probability Tables (HD-CPT). When a sensor fails, the sensor is managed in that condition in the corresponding MSNBC2s. For instance, if the PRE sensor is not available, there are no more data from it, so the data of the other working sensors should supply information to recover the behaviour of the missing sensor, to limit the reduction in the recognition performance of the CCC-FSNBC. This example is from the post-stroke patient rehabilitation dataset [5], [7], which includes PRE, MOV, and FAE sensors; and the affective states of tiredness, anxiety, pain, and engagement. In this example, MSNBC2 operationalize five odd-size sliding windows W, $|W| = 3, 5, 7, 9, 11$ Acronyms meanings: C_j = class variable (a categorical affective state), $j \in \{1, 2, \ldots, q\}$, q = number of affective states (in this example, q = 4). $C_j^{\prime}, j \in \{1, 2, \ldots, q-1\}$ = predicted classes. C_{sjh} = class variable for the inference of MSNBC2 for sensor sjh, $h \in \{1, 2, \ldots, m\}$, m = number of sensors (in this example, m = 3).

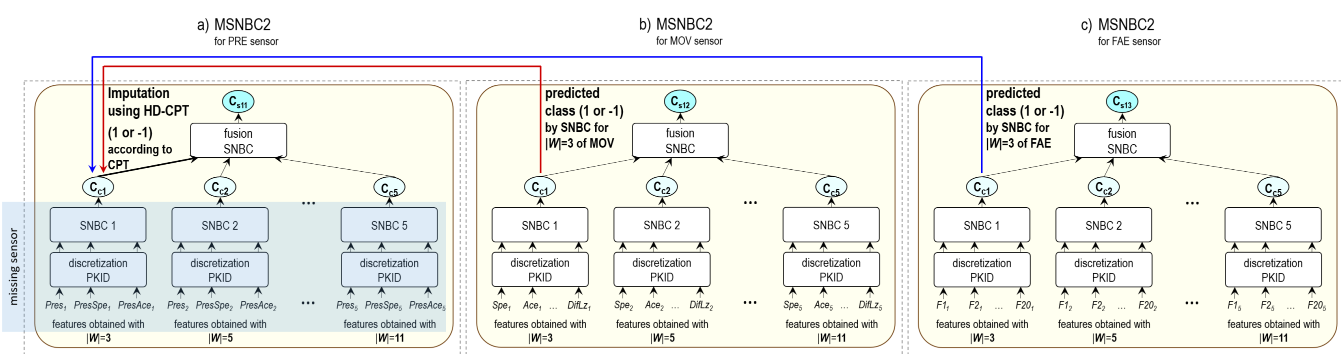

Fig. 4. Imputation process of Hot Deck using Conditional Probability Tables (HD-CPT) for the class of the respective odd-size sliding window $|W| = 3, 5, 7, 9, 11$ of MSNBC2 of the missing sensor. Exemplification corresponds to the dataset of the rehabilitation of post-stroke patients [5], [7], which includes the sensors of PRE, MOV, and FAE. Since the input features are numeric values, a discretization process called Proportional k-interval discretization (PKID) [22] was used. In the example, the PRE sensor fails, and the other sensors, MOV, and FAE provide the information to find the donor vectors in the historical records. The imputation is indicated for the window $|W| = 3$, but the process is similar for the other odd-size windows $|W| = 5, 7, 9, 11$. The imputation can be done by selecting the class 1 or -1 according to the probability distribution registered in the Conditional Probability Table (CPT) of the classes of the receiver given the classes of the donors.

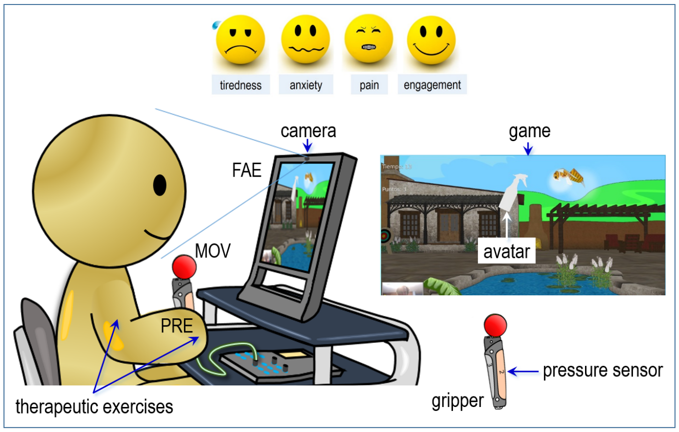

Fig. 5. Schematic depiction of a person interacting with the virtual rehabilitation platform, Gesture Therapy. The person is holding the gripper, which has a frontal sensor for registering finger pressure. The webcam follows the gripper's colour ball to control an avatar (in this case is the insecticide bottle on the screen) in the game "garden pond" (virtual environment).

(Fig. 5). The tracker system estimates the 3D coordinates of hand movements (MOV sensor) and the finger pressure (PRE sensor) value at each video frame. GT also includes capabilities for video recording so that the patient's upper torso and spontaneous facial expressions (FAcial Expression: FAE sensor) can be captured while playing the games.

VI. Dataset

To illustrate the efficacy of our method, we apply it in neurorehabilitation. A dataset of post-stroke patients [5], [7] undergoing neurorehabilitation as administered through GT was used to assess the proposed method, HD-CPT, for dealing with a missing sensor. The dataset contains the records of 5 post-stroke patients who performed therapeutic exercises using the GT platform during ten longitudinal sessions over a period of about one month (each session was taken on a different day, maximum 3 sessions per week.). Frontal videos of the patients were recorded while playing the games. Meanwhile, data were collected at each video frame and consisted of :

1) a finger pressure value (from PRE sensor),
2) a 3D hand position (from MOV sensor), and
3) a facial expression (from FAE sensor).

Data were labelled frame by frame by psychiatrists using four binary values indicating the presence (1) or the absence (-1) of each of the four states: tiredness, anxiety, pain, and engagement. Fleiss' κ suggested substantial agreement for tiredness, moderate agreement for engagement and fair agreement for anxiety and pain [7]. More than one state could be present at the same time (multilabel classification scheme).

Feature vectors were generated with a sliding window W, of a predefined odd size $|W| = 3, 5, 7, 9, 11$, over consecutive frames of the respective sensor data. For each step forward of the sliding window, a new feature vector was generated, which has the following components (each component is an *average of* the respective calculation in the sliding window):

PRE 3 features: *pressure* (*Pres*), *pressure speed* (*PresSpe*) and *pressure acceleration* (*PresAce*);

MOV 5 features: *speed* (*Spe*), *acceleration* (*Ace*) and *differential location* by the axes: x (*DifLx*), y (*DifLy*), z (*DifLz*); and

FAE 20 features from each frame of the patients' frontal video [24]. These features represent average distances or angles of geometrical figures over the eyebrows, the eyes, and the mouth [25]. Then, the feature vector contains 20 averaged values over the sliding window.

All the feature vectors have four binary tags (from the set $\{-1, 1\}$), one for each state (tiredness, anxiety, pain, and engagement), representing the presence (1) or the absence (-1) of the state. These tags were generated considering a sliding window (synchronized through the frames with the sliding windows of the feature vectors) and the majority label in the sliding window was selected as the corresponding tag.

VII. Experiments and Results

For evaluating the performance of the proposed method, HD-CPT, incorporated into the multiresolution classifier MSNBC2, we consider a missing sensor at a time and analyse the corresponding results. There are three scenarios: (a) the model where only the available sensors are considered, i.e. where there is no influence of the missing sensor, (b) the model where the missing sensor's class is estimated according to the proposed strategy, and (c) the model with all sensors available. The results of the three models were contrasted to see if the model of scenario (b) achieved a better performance than the model of scenario (a) and show a close performance to scenario (c).

The information of the missing sensor was estimated through the HD-CPT method employing the information of the remaining available sensors. Therefore, FSNBC received two types of inputs from the MSNBC2s: inputs obtained from the available sensors and inputs generated over estimations for the missing sensor through HD-CPT. Experiments in three situations were carried out to compare the performance of the three models: (i) when PRE sensor fails, (ii) when MOV sensor fails, and (iii) when FAE sensor fails.

CCC-FSNBC models were independently built for each patient (in a within-subject setting for studying the system performance for the customization to each patient) to predict the occurrence of the four states (tiredness, anxiety, pain, and engagement) in the multilabel classification scheme. Therefore, we had 5 CCC models, one for each patient. For each CCC, there were as many FSNBCs models as affective states, so there were 4 FSNBCs. The corresponding MSNBC2s were constructed using the sliding windows of odd-sizes 3 to 11: $|W| = 3, 5, 7, 9, 11$; so 5 different window sizes were used.

Stratified 10-fold cross-validation across all the rehabilitation sessions was applied for internal validity. The performance of the respective computational models, for each patient and for the three different sensor scenarios, was evaluated using several metrics for multilabel classification [26]: Global accuracy ($GAcc$), Mean accuracy ($MAcc$), Multi-label accuracy ($MLAcc$), and $F\text{-}measure$.

The results of the proposed method HD-CPT when a sensor fails: PRE, MOV, or FAE are presented in Table I, Table II, and Table III, respectively, and the results are summarized as $mean \pm std.\,deviation$ across the 5 patients and across the 10 folds of the cross-validation. Concerning the training process, the CCC model was run with 8 iterations in each experiment. The results are presented in the tables considering the following order: First, the results of scenario (a), where one sensor is absent and the other two remaining sensors are only considered. Second, the results of scenario (b), where the missing sensor has been treated according to the proposed method HD-CPT. And third, the results of scenario (c), where all the sensors are available.

A. Performance comparison when the PRE sensor fails

Table I presents the results of the proposed method HD-CPT when the PRE sensor fails, called $\widehat{PRE}$-MOV-FAE. In all cases, the results of CCC-FSNBC for $\widehat{PRE}$-MOV-FAE are between the results of scenario (a) and the results of all the sensors available –scenario (c)–, but there are no significant differences for $GAcc$ (Friedman test: $\chi^2(2) = 5.709$, $p = 0.058$), $MAcc$ (Friedman test: $\chi^2(2) = 4.508$, $p = 0.105$), $MLAcc$ (Friedman test: $\chi^2(2) = 5.992$, $p = 0.050$), and $F-measure$ (Friedman test: $\chi^2(2) = 5.992$, $p = 0.050$).

TABLE I

PERFORMANCE COMPARISON ACROSS THE 5 PATIENTS AND THE 10 FOLDS OF CROSS-VALIDATION ($mean \pm std.dev.$) WHEN THE PRE SENSOR FAILS.

Sensors	$GAcc$	$MAcc$	$MLAcc$	$F-measure$
MOV-FAE	0.935 ± 0.066	0.975 ± 0.025	0.948 ± 0.053	0.952 ± 0.049
$\widehat{PRE}$-MOV-FAE	0.936 ± 0.064	0.976 ± 0.024	0.949 ± 0.052	0.953 ± 0.049
PRE-MOV-FAE	0.941 ± 0.059	0.977 ± 0.023	0.954 ± 0.046	0.958 ± 0.043

B. Performance comparison when the MOV sensor fails

Table II presents the results of the proposed method HD-CPT when the MOV sensor fails, called PRE-$\widehat{MOV}$-FAE. In all cases, the results of CCC-FSNBC for PRE-$\widehat{MOV}$-FAE are between the results of scenario (a) and the results of all the sensors available –scenario (c)–. The results of CCC-FSNBC for PRE-MOV-FAE are significantly higher than the ones for PRE-FAE and PRE-$\widehat{MOV}$-FAE (Friedman test, $p < 0.05$, with post hoc analysis with Wilcoxon signed-rank tests with Bonferroni correction, $p < 0.017$).

C. Performance comparison when the FAE sensor fails

Table III presents the results of the proposed method HD-CPT when the FAE sensor fails, called PRE-MOV-$\widehat{FAE}$. In all cases, the results of CCC-FSNBC for PRE-MOV-$\widehat{FAE}$ are between the results of scenario (a) and the results of all the sensors available –scenario (c)–. The results of CCC-FSNBC for PRE-MOV-FAE are significantly higher than the ones for PRE-MOV and PRE-MOV-$\widehat{FAE}$ (Friedman test, $p < 0.05$, with post hoc analysis with Wilcoxon signed-rank tests with Bonferroni correction, $p < 0.017$).

TABLE II

PERFORMANCE COMPARISON ACROSS THE 5 PATIENTS AND THE 10 FOLDS OF CROSS-VALIDATION ($mean \pm std.dev.$) WHEN THE MOV SENSOR FAILS.

Sensors	$GAcc$	$MAcc$	$MLAcc$	$F-measure$
PRE-FAE	0.915 ± 0.085	0.967 ± 0.033	0.929 ± 0.073	0.934 ± 0.069
PRE-$\widehat{MOV}$-FAE	0.924 ± 0.076	0.972 ± 0.028 †	0.938 ± 0.063	0.943 ± 0.060
PRE-MOV-FAE	0.941 ± 0.059 ‡	0.977 ± 0.023 ‡	0.954 ± 0.046 ‡	0.958 ± 0.043 ‡

‡ means significant differences between CCC of PRE-MOV-FAE and CCC of PRE-FAE, and between CCC of PRE-MOV-FAE and CCC of PRE-$\widehat{MOV}$-FAE (Friedman test, for $GAcc$ $\chi^2(2) = 25.861$, $p < 0.05$, for $MAcc$ $\chi^2(2) = 28.203, p < 0.05$, for $MLAcc$ $\chi^2(2) = 24.356, p < 0.05$, and for $F - measure$ $\chi^2(2) = 22.946, p < 0.05$, post hoc analysis with Wilcoxon signed-rank tests with Bonferroni correction, $p < 0.017$).

† means significant differences between CCC of PRE-$\widehat{MOV}$-FAE and CCC of PRE-FAE $W = -2.851, p < 0.017$

TABLE III

PERFORMANCE COMPARISON ACROSS THE 5 PATIENTS AND THE 10 FOLDS OF CROSS-VALIDATION ($mean \pm std.dev.$) WHEN THE FAE SENSOR FAILS.

Sensors	$GAcc$	$MAcc$	$MLAcc$	$F-measure$
PRE-MOV	0.851 ± 0.117	0.939 ± 0.045	0.875 ± 0.098	0.883 ± 0.093
PRE-MOV-$\widehat{FAE}$	0.860 ± 0.099	0.944 ± 0.040	0.885 ± 0.084	0.893 ± 0.081
PRE-MOV-FAE	0.941 ± 0.059 ‡	0.977 ± 0.023 ‡	0.954 ± 0.046 ‡	0.958 ± 0.043 ‡

‡ means significant differences between CCC of PRE-MOV-FAE and CCC of PRE-MOV, and between CCC of PRE-MOV-FAE and CCC of PRE-MOV-$\widehat{FAE}$ (Friedman test, for $GAcc$ $\chi^2(2) = 50.839$, $p < 0.05$, for $MAcc$ $\chi^2(2) = 50.667, p < 0.05$, for $MLAcc$ $\chi^2(2) = 50.559, p < 0.05$, and for $F - measure$ $\chi^2(2) = 50.559, p < 0.05$, post hoc analysis with Wilcoxon signed-rank tests with Bonferroni correction, $p < 0.017$).

D. Training Times

The average training time of CCC-FSNBC in scenario (b), where a sensor fails, $\widehat{PRE}$, $\widehat{MOV}$, or $\widehat{FAE}$ was ($mean \pm std.\,deviation$) 125.27 ± 27.82 sec., or 2.09 ± 0.46 min. The average training time of CCC-FSNBC for PRE-MOV-FAE –scenario (c)– was ($mean \pm std.\,deviation$) 124.09 ± 30.46 sec. which corresponds to 2.07 ± 0.51 min. Therefore, the proposed method HD-CPT represented an average increase of 1.18 sec., approximately, in the average training time of CCC-FSNBC.

VIII. DISCUSSION

The problem of the absence of a sensor at the testing phase has been explored with a strategy of imputation, which estimates the values of the missing sensor through the HD-CPT employing the information of the remaining available sensors. According to the results, the performance of the chosen strategy in a three-sensor problem was generally successful, as the results fall between the use of all sensors and the baseline of just eliminating one sensor. In particular, for the MOV and FAE sensors, the estimated values contributed to improve the performance of CCC-FSNBC with respect to scenario (a); while for the PRE sensor the results were still better than scenario (a), although the difference is lower. This could be because the PRE sensor is the one that contributes more to the recognition process. Therefore, these results suggest that when a specific sensor fails, its predicted classes can be estimated through the predicted classes of the remaining sensors using the proposed strategy, but the results will depend on the specific sensor. Moreover, the conjunction of information from different sensors can contribute to the recognition of affective states, but this is influenced by relations of complementarity, redundancy, or noise between sensors. With respect to the training process, HD-CPT increased the execution time by 1.18 sec., approximately in relation to the model of all sensors available. So the proposed method does not represent an overhead for the computational model.

IX. Conclusions and Future Work

We have proposed a novel method, HD-CPT, for dealing with the loss of a sensor in a multilabel and multimodal affective states recognition system. We have exemplified the use of the method in an existing platform for the recognition of the affective states of patients during physical rehabilitation. The proposed method was beneficial when one sensor failed, achieving higher results in general than when not using such mechanism (scenario (a)). HD-CPT does not add extra computational cost when a sensor is lost, which is an advantage over other methods and gives robustness, preventing the system from collapsing; it provides an alternative for the use of automatic affective states recognition systems in naturalistic everyday life. An issue to consider is which sensor fails and which sensor is absolutely necessary, i.e., if the decrease in performance due to the estimated values of a missing sensor may be acceptable.

As future work, the problem of a missing sensor in systems of affective computing should be studied with other sensors as the ones from signals of EEG, functional Near-Infrared Spectroscopy (fNIRS), ECG, and GSR. A larger trial is necessary to confirm whether this apparent trend can be generalized to the population considered. The proposed method should be studied for the case when more than one sensor fails.

References

[1] G. Kalton and D. Kasprzyk, "The treatment of missing survey data," *Survey methodology*, vol. 12, pp. 1–16, 1986.

[2] R. R. Andridge and R. J. Little, "A review of hot deck imputation for survey non-response," *International Statistical Review*, vol. 78, no. 1, pp. 40–64, 2010.

[3] J. J. Rivas, F. Orihuela-Espina, and L. E. Succar, "Circular chain classifiers," in *Proceedings of Machine Learning Research. 9th International Conference on Probabilistic Graphical Models*, vol. 72, 2018, pp. 380–391.

[4] J. J. Rivas, F. Orihuela-Espina, L. E. Sucar, A. Williams, and N. Bianchi-Berthouze, "Automatic recognition of multiple affective states in virtual rehabilitation by exploiting the dependency relationships," in *Proceedings of the IEEE 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2019, pp. 1–7.

[5] J. J. Rivas, M. d. C. Lara, L. Castrejón, J. Hernández-Franco, F. Orihuela-Espina, L. Palafox, A. Williams, N. Bianchi-Berthouze, and L. E. Sucar, "Multi-label and multimodal classifier for affective states recognition in virtual rehabilitation," *IEEE Transactions on Affective Computing*, 2021.

[6] J. J. Rivas, F. Orihuela-Espina, and L. E. Sucar, "Recognition of affective states in virtual rehabilitation using late fusion with Semi-Naive Bayesian classifier," in *Proceedings of the ACM 13th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth 2019)*, 2019, pp. 308–313.

[7] J. J. Rivas, F. Orihuela-Espina, L. Palafox, N. Berthouze, M. del Carmen Lara, J. Hernández-Franco, and E. Sucar, "Unobtrusive inference of affective states in virtual rehabilitation from upper limb motions: A feasibility study," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 470–481, 2018.

[8] L. E. Sucar, R. Leder, J. Hernández, I. Sánchez, and G. Azcárate, "Clinical evaluation of a low-cost alternative for stroke rehabilitation," in *Proceedings of the IEEE 11th International Conference on Rehabilitation Robotics (ICORR 2009)*, 2009, pp. 863–866.

[9] L. E. Sucar, F. Orihuela-Espina, R. L. Velazquez, D. J. Reinkensmeyer, R. Leder, and J. Hernández-Franco, "Gesture therapy: An upper limb virtual reality-based motor rehabilitation platform," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 3, pp. 634–643, 2014.

[10] V. Vapnik and A. Vashist, "A new learning paradigm: Learning using privileged information," *Neural Networks*, vol. 22, no. 5, pp. 544–557, 2009.

[11] Z. Wang, X. Wang, and Q. Ji, "Learning with hidden information," in *Proceedings of the IEEE 22nd International Conference on Pattern Recognition (ICPR 2014)*, 2014, pp. 238–243.

[12] Z. Wang and Q. Ji, "Classifier learning with hidden information," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4969–4977.

[13] M. Soleymani and M. Pantic, "Human-centered implicit tagging: Overview and perspectives," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC 2012)*, 2012, pp. 3304–3309.

[14] S. Wang, Y. Zhu, L. Yue, and Q. Ji, "Emotion recognition with the help of privileged information," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 189–200, 2015.

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

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

[17] Z. Gao and S. Wang, "Emotion recognition from EEG signals using hierarchical Bayesian network with privileged information," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, 2015, pp. 579–582.

[18] S. Wang, M. He, Y. Zhu, S. He, Y. Liu, and Q. Ji, "Learning with privileged information using Bayesian networks," *Frontiers of Computer Science*, vol. 9, no. 2, pp. 185–199, 2015.

[19] R. O. Duda, P. E. Hart *et al.*, *Pattern classification and scene analysis*. Wiley New York, 1973, vol. 3.

[20] M. J. Pazzani, "Searching for dependencies in Bayesian classifiers," in *Learning from Data*. Springer, 1996, pp. 239–248.

[21] J. J. Rivas, "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), Puebla, México., 2015.

[22] Y. Yang and G. I. Webb, "Proportional k-interval discretization for naive-bayes classifiers," in *Proceedings of the 12th European Conference on Machine Learning (ECML 2001)*. Springer, 2001, pp. 564–575.

[23] J. J. Rivas, F. Orihuela-Espina, L. E. Sucar, L. Palafox, J. Hernández-Franco, and N. Bianchi-Berthouze, "Detecting affective states in virtual rehabilitation," in *Proceedings of the IEEE 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth 2015)*, 2015, pp. 287–292.

[24] A. Bandini, S. Orlandi, H. J. Escalante, F. Giovannelli, M. Cincotta, C. A. Reyes-Garcia, P. Vanni, G. Zaccara, and C. Manfredi, "Analysis of facial expressions in parkinson's disease through video-based automatic methods," *Journal of neuroscience methods*, vol. 281, pp. 7–20, 2017.

[25] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42–55, 2012.

[26] C. Bielza, G. Li, and P. Larrañaga, "Multi-dimensional classification with Bayesian networks," *International Journal of Approximate Reasoning*, vol. 52, no. 6, pp. 705–727, 2011.