Seven principles to mine flexible behavior from physiological signals for effective emotion recognition and description in affective interactions

Rui Henriques[1] and Ana Paiva[2]

[1]*KDBIO, Inesc-ID, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal*

[2]*GAIPS, Inesc-ID, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal*

{*rmch,ana.s.paiva*}*@ist.utl.pt*

Keywords: Mining Physiological Signals, Measuring Affective Interactions, Emotion Recognition and Description

Abstract: Measuring affective interactions using physiological signals has become a critical step to understand engagements with human and artificial agents. However, traditional methods for signal analysis are not yet able to effectively deal with the differences of responses across individuals and with flexible sequential behavior. In this work, we rely on empirical results to define seven principles for a robust mining of physiological signals to recognize and characterize affective states. The majority of these principles are novel and driven from advanced pre-processing techniques and temporal data mining methods. A methodology that integrates these principles is proposed and validated using electrodermal signals collected during human-to-human and human-to-robot affective interactions.

1 Introduction

Monitoring physiological signals is increasingly necessary to derive accurate analysis from affective interactions or to dynamically adapt these interactions. Although many methods have been proposed for an emotion-centered analysis of physiological signals (Jerritta et al., 2011; Wagner et al., 2005), there is still lacking an integrative view of existing contributions. Additionally, existing methods suffer from three major drawbacks. First, there is no agreement on how to deal with individual differences and with spontaneous variations of the signals. Second, they generally rely on feature-driven models and, therefore, discard flexible sequential behavior of physiological responses. Finally, experimental conditions and psychophysiological data from users have not been adopted to shape the classification models.

In this paper, we propose seven principles to guide the mining of physiological signals for an effective emotion recognition and characterization. These principles were derived from an experimental comparison of advanced techniques from machine learning and signal processing using physiological signals, such as skin activity and temperature, collected during affective interactions. These principles can be used to address the three introduced drawbacks. They provide an integrated and up-to-date view on how to disclose and describe affective states from physiological signals. A methodology that relies on these principles is, additionally, proposed.

This paper is structured as follows. In *Section 2*, relevant work on the mining of sensor-based data in emotion-centered studies is covered. *Section 3* defines the seven principles and the target methodology. *Section 4* provides the supporting quantitative evidence for the introduced principles using signals collected under different experimental settings. Finally, the main implications are synthesized.

2 Background

Physiological signals are increasingly adopted to monitor and shape affective interactions since they are hardly prone to masking and can track subtle but significant cognitive-sensitive emotional changes. However, their complex, variable and subjective expression within and among individuals pose key challenges for an adequate modeling of emotions.

Consider a set of annotated signals D=$(x_1,..,x_m)$, where each instance is a tuple x_i=$(\vec{y},a_1,..,a_n,c)$ where $\vec{y}$ is the signal, a_i is an annotation related with the subject or experimental setting, and c is the labeled emotion or stimulus. Given D, the *emotion recognition* task aims learn a model M to label a new unlabeled instance $(\vec{y},a_1,..,a_n)$. *Emotion description* task aims to learn a model M that characterizes the major properties of $\vec{y}$ signal for each emotion c.

The goal of emotion recognition and description is to (dynamically) access someone's feelings from (streaming) signals. Emotion recognition from physiological signals has been applied in the context of human-robot interaction (Kulic and Croft, 2007; Leite et al., 2013), human-computer interaction (Picard et al., 2001), social interaction (Wagner et al., 2005), sophisticated virtual adaptive scenarios (Rani et al., 2006), among others (Jerritta et al., 2011). Multiple physiological modalities have been adopted depending on the goal of the task. For instance, electrodermal activity has been used to identify engagement and excitement states, respiratory volume and rate to recognize negative-valenced emotions, and heat contractile activity to separate positive-valenced emotions (Wu et al., 2011). Additionally, the experimental setting of existing studies also vary, namely the properties of the selected stimuli (discrete vs. continuous) and general factors related with user dependency (single vs. multiple subjects), subjectivity of the stimuli (high-agreement vs. self-report) and the analysis time of the signal (static vs. dynamic).

A first drawback of existing emotion-centered studies is the absence of learned principles to mine the signals. Although multiple models are compared using accuracy levels, there is no in-depth analysis of the underlying behavior of these models and no guarantees regarding their statistical significance. Additionally, there is no assessment on how their performance varies for alternative experimental settings.

A second drawback is related with the fact that these studies rely on simple pre-processing techniques and feature-driven models. First, preprocessing steps are centered on the removal of contaminations and on simplistic normalization procedures. These techniques are insufficient to deal with differences on responses among subjects and with the isolation of spontaneous variations of the signal.

Second, even in the presence of expressive features, models are not able to effectively accommodate flexible sequential behavior. For instance, a rising or recovering behavior may be described by specific motifs sensitive to sub-peaks or displaying a logarithmic decaying. This weak-differentiation among responses leads to rigid models of emotions.

The task of this work is to identify a set of consistent principles to address these drawbacks, thus improving emotion recognition rates.

3 Solution

Relying on experimental evidence, seven principles were defined to surpass the limitations of traditional models for emotion recognition from physiological signals. The impact of adopting these principles were validated over electrodermal activity, facial expression and skin temperature signals. Nevertheless, these principles can be tested for any other physiological signal after the neutralization of cyclic behavior (e.g. respiratory and cardiac signals) and/or the application of smoothing and low-pass filters.

3.1 The Seven Principles

#1: Adopt representations able to handle individual differences of responses

Problem: The differences of physiological responses for a single emotion are often related with experimental conditions, such as the placement of sensors or unregulated environment, and with specific psychophysiological properties of the subjects, such as lability and current mood. These undesirable differences affect both the: *i)* amplitude axis (varying baseline levels and peak-variations of responses), and the *ii)* temporal axis (varying latency, rising and recovery time of responses).

On one hand, recognition rates degrade as a result of an increased modeling complexity due to these differences. On the other hand, when normalizing signals along the amplitude-time axes, we are discarding absolute behavior that is often critical to distinguish emotions. Additionally, common normalization procedures are not adequate since the signal baseline and response amplitude may not be correlated (e.g. high baseline does not mean heightened elicited responses).

Solution: A new representation of the signal that minimizes individual differences should be adopted, and combined with the original signal for the learning of the target model.

While many representations for time series exist (Lin et al., 2003b), they either scale poorly as the cardinality is not changed or require previous access to all the signal preventing a dynamic analysis of the signal. Symbolic ApproXimation (SAX) satisfies these requirements and offers a lower-bounding guarantee. SAX behavior can be synthesized in two steps. First, the signal is transformed into a Piecewise Aggregate Approximated (PAA) representation. Second, the PAA signal is symbolized into a discrete string. A Gaussian distribution is used to produce symbols with equiprobability from statistical breakpoints (Lin et al., 2003a). Unlike other representations, the Gaussian distribution for amplitude control smooths the problem of subjects with baseline levels and response variations not correlated.

Amplitude differences can be corrected with respect to all stimuli, to a target stimulus, to all subjects, or to a specific subject. To treat temporal differences, two strategies can be adopted. First, signals can be used as-is (with their different numerosity) and given as input to sequential learners, which are able to deal with this aspect. Note, for instance, the robustness of hidden Markov models on detecting hand-writing text with different sizes in (Bishop, 2006). Second, the use of piecewise aggregation analysis, such as provided by SAX, can be used to normalize numerosity differences.

#2: Account for relevant signal variations

Problem: Motifs and features sensitive to sub-peaks are critical for emotion recognition (e.g. electrodermal variations hold the potential to separate anger from fear responses (Andreassi, 2007)). However, traditional methods rely on fixed amplitude-thresholds to detect informative signal variations, which became easily corrupted due to the individual subject differences. Additionally, when cardinality is reduced, relevant sub-peaks disappear.

Solution: Two strategies can be adopted. First, a representation to enhance local variations, referred as local-angle. The signal is partitioned in thin time-partitions and the angle associated with the signal variation for each partition is computed and translated into symbols based on break-points computed from the input number of symbols. Similarly to SAX, the angle break points are also defined assuming a Gaussian distribution. When adopting an 6-dim alphabet, the following illustrative SAX-based univariate signal: <17,13,15,14,18,19,16,14,13,12,16,16>, would be translated into the following local-angle representation: <0,4,1,5,5,0,1,1,1,5,4>.

Second, multiple SAX representations can be adopted using different cardinalities. While mapping the raw signals into low-cardinal signals is useful to capture smoothed behavior (e.g. alphabet size less than 8), a map into high-cardinal signals is able to capture more delineated behavior (e.g. alphabet size above 10). One model can be learned for each representation, with the joint probability being computed to label a response.

#3: Include flexible sequential behavior

Problem: Although sequential learning is the natural option for audio-and-visual signals, the existing models for emotion recognition mainly rely on extracted features. Feature-extraction methods are not able to capture flexible behavior (e.g. motifs underlying complex rising and decaying responses) and are strongly dependent on directive thresholds (e.g. peak amplitude to compute frequency measures).

Solution: Generative models learned from sequential data, such as recurrent neural networks or dynamic Bayesian networks, can be adopted to satisfy this principle (Bishop, 2006). In particular, hidden Markov models (HMMs) are an attractive option due to their stability, simplicity and flexible parameter-control (Murphy, 2002). The core task is to learn the generation and transition probabilities of a hidden automaton for each emotion. Given a non-labeled signal, we can assess the probability of being generated by each learned model. An additional exploitation of the lattices per emotion can be used to retrieve emerging patterns and, thus, be used as emotion descriptors.

The parameterization of HMMs must be based on the signal properties (e.g. high dimensionality leads to an increased number of hidden states). Alternative architectures, such as fully-interconnected or left-to-right architectures, can be considered.

From the conducted experiments, an analysis of the learned emissions from the main path of left-to-right HMM architectures revealed emerging rising and recovering responses following sequential patterns with flexible displays (e.g. exponential and "stairs"-appearance behavior).

#4: Integrate sequential and feature-driven models

Problem: Since sequential learners capture the overall behavior of physiological responses, they are not able to highlight specific discriminative properties of the signal. Often such discriminative properties are adequately described by simple features.

Solution: Feature-driven and sequential models should be integrated as they provide different but complementary views. One option is to rely on a post-voting stage. A second option is to use one model to discriminate the less probable emotions, and to use such constraints on the remaining model.

Feature-driven models have been widely researched and are centered on three major steps: feature extraction, feature selection and feature-based learning (Lessard, 2006; Jerritta et al., 2011). Expressive features include statistical, temporal, frequency and, more interesting, temporal-frequency metrics (from geometric analysis, multiscale sample entropy, sub-band spectra). Feature extraction methods include tonic-phasic windows; moving-sliding features; transformations (Fourier, wavelet, Hilbert); component analysis; projection pursuit; auto-associative nets; and self-organizing maps. Methods to remove features without significant correlation with the emotion under assessment include sequential selection, branch-and-bound search, Fisher projection, Davies-Bouldin index, analysis of variance and some classifiers. Finally, a wide-variety of deterministic and probabilistic learners have been adopted to perform emotion recognition based on relevant features.

The most successful learners are k-nearest neighbors, regression trees, random forests, Bayesian networks, support vector machines, canonical correlation and linear discriminant analysis, neural networks, and Marquardt-back propagation.

#5: Use subject's traits to shape the model

Problem: Subjects with different psychophysiological profiles tend to have different physiological responses for the same stimuli. Modeling responses for emotions without this prior knowledge hampers the learning task since the models have to define multiple paths or generalize responses in order to accommodate such alternative expressions of an emotion due to profile differences.

Solution: Turn the learning sensitive to psychophysiological traits of the subject under assessment when available. We found that the inclusion of the relative score for the four Myers-Briggs types[1] was found to increase the accuracy of learning models.

For lazy learners, the simple inclusion of these traits as features is sufficient. We observed an increased accuracy in k-nearest neighbors, which tends to select responses from subjects with related profile.

A simple strategy for non-lazy learners is to partition data by traits, and to learn one model for each trait. Emotion recognition is done by integrating the results of the models with the profile of the testing subject. This integration can recur to a weighted voting scheme, where weights essentially depend on the score obtained for each assessed trait.

A more robust strategy is to learn a tree structure with classification models in the leafs, where a branching decision is associated with trait values that are correlated with heightened response differences for a specific emotion.

#6: Refine the learning models based on the complexity of emotion expression

Problem: A single emotion-evocative stimulus can elicit small-to-large groups of significantly different physiological responses. A simple generalization of each set of responses leads to poor models.

Solution: Create multiple sub-models for emotions with varying physiological expressions. Both rule-based models, such as random forests, and lazy learners implicitly accommodate this behavior.

Generative models need to be further refined when the emission probabilities of the underlying lattices for a specific emotion do not have a strong convergence. When HMMs are adopted, it is crucial to change the architecture to add an alternative path with a new hidden automaton.

For non-generative models, it is crucial to understand when the model needs to be further refined. This can be done by analyzing the variances of features per emotion or by clustering responses per emotion with a non-fixed number of clusters.

Not only these strategies can improve the emotion recognition rates, but also the characterization of physiological responses per emotion. Consider the case where the learned HMMs are used as a pattern descriptor. Without further separation of different expressions for each emotion, the generative models per emotion would be more prone to error and only reveal generic behavior.

#7: Affect the models to the conditions of the experimental setting

Problem: the properties of the emotion recognition task varies with different settings, such as discrete vs. prolonged stimuli, user-dependent vs. independent studies, univariate vs. multivariate signals.

Solution: The selection and parameterization of classification models should be guided by the experimental conditions. Below we introduced three examples derived from our analysis. First, the influence of sub-peak analysis (principle #2) for emotion recognition should have a higher weight for prolonged stimuli. Second, user-dependent studies are particularly well described by flexible sequential behavior (principle #3). Third, multivariate analysis should be performed in an integrated fashion whenever possible. Common generative models, such as HMMs, are able to model multivariate signals.

Additionally, we found that both the inclusion of other experimental properties (such as interaction annotations) and of the perception of the subject regarding the interaction (assessed recurring to post-surveys) can guide the learning of the target emotion recognition models.

3.2 Methodology

Relying on the introduced seven principles, we propose a novel methodology for emotion recognition and description from physiological signals[2]. Fig.1 illustrates its main steps. Emotion recognition combines the traditional feature-based classification with the results provided from sequence learners and is centered on two expressive representations: *i)* SAX to normalize individual differences while still preserving overall response pattern, and on *ii)* local angles to enhance the local sub-peaks of a response. Additionally, emotion characterization is accomplished using both feature-based descriptors (mean and variance

[1] http://www.myersbriggs.org/

[2] Software in web.ist.utl.pt/rmch/research/software/eda

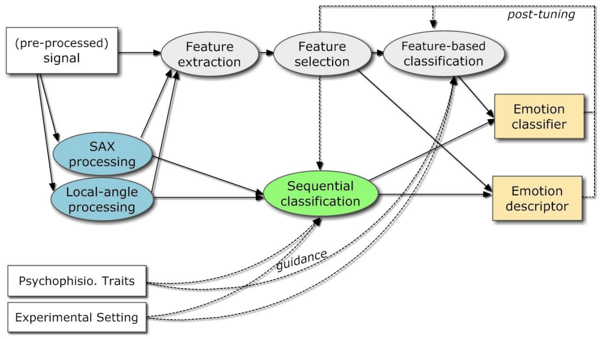

Figure 1: Proposed methodology for emotion recognition and description from physiological signals

of the most discriminative features) and the transition lattices generated by sequence learners.

In the presence of background knowledge, that is, when each instance $(\vec{y}, a_1, .., a_n, c)$ has $n \geq 1$, prior decisions can be made. Exemplifying, in the presence of psychophysiological traits correlated with varying expression of a specific emotion, the target model can be further decomposed to reduce the complexity of the task. Complementary, iterative refinements over the learned model can be made when feature-based models rely on features with high variances or when the generative models do not have strong convergence criteria for a specific emotion.

4 Results

The proposed principles and methodology resulted from an evaluation of advanced data mining and signal processing concepts using a tightly-controlled lab study[3]. More than 200 signals were collected for each physiological modality from both human-to-human and human-to-robot affective interactions[4]. Electrodermal activity (EDA), skin temperature, and facial expression modalities were monitored using Affectiva technology. Although the conveyed results are centered on electrodermal activity and temperature, previous work from the institute on the use of facial expression to recognize emotion during affective games adds supporting evidence to the relevance of the listed principles (Leite et al., 2013).

Eight different stimuli, 5 emotion-centered stimuli[5] and 3 others (captured during periods of strong physical effort, concentration and resting), were presented to each subject[6]. A survey was used to categorize the profile of the participants according to the Myers-Briggs type indicator.

Statistical and geometric features were extracted from the raw, SAX and local-angle representations. Feature selection was performed using statistical analysis of variance (ANOVA). The selected feature-based classifiers were adopted from WEKA software (Hall et al., 2009), and the HMMs from HMM-WEKA extension (codified according to Bishop (2006)). SAX and local angle representations were implemented using Java (JVM version 1.6.0-24) and the following results were computed using an Intel Core i5 2.80GHz with 6GB of RAM.

Principles #1 and #2. To assess the impact of dealing with individual differences and informative subtle variations of the signals, we evaluate emotion recognition scores under SAX and local-angle representations using feature-driven models. The score is accuracy, the ability to correctly label an unlabeled signal (i.e. to identify the underline emotion from 5 emotions). Accuracy was computed using a 10 cross-fold validation over the ~200 collected electrodermal signals. Fig.2 synthesizes the results.

The isolated use of electrodermal features from the raw signal (tonic and phasic skin conductivity, maximum amplitude, rising and recovering time) and of statistical features extracted from SAX and local-angle representations leads to an accuracy near 50% (against 20% when using a random model). The integration of these features results in an improvement of 10pp to near 60%. Additionally, accuracy improves when features from skin temperature are included.

Logistic learners, which use regressions on the real-valued features to affect the probability score of each emotion, were the best feature-based models for this experiment. When no feature selection method is applied, Bayesian nets are an attractive alternative. Despite the differences between human-to-human and human-to-robot settings, classifiers are still able to recognize emotions when mixing the cases. For in-

[3]details, data, scripts and statistical sheets available in http://web.ist.utl.pt/rmch/research/software/eda

[4]30 participants, with ages between 19 and 24 (average of 21 years old), were randomly divided in two groups, R and H. Subjects from group R interacted with the NAO robot (www.aldebaran-robotics.com) using a wizard-of-Oz setting. Participants from group H interacted with an human agent, an actor with a structured and flexible script.

[5]Empathy (following common practices in speech tone and body approach), expectation (possibility of gaining an additional reward), positive-surprise (unexpected attribution of a significant incremental reward), stress (impossible riddle to solve in a short time to maintain the incremental reward) and frustration (self-responsible loss of the initial and incremental rewards).

[6]The stimuli were presented in the same order in every experience and 6-8 minutes was provided between two stimulus to neutralize the subject emotional state and remove the stress related with the experimental expectations.

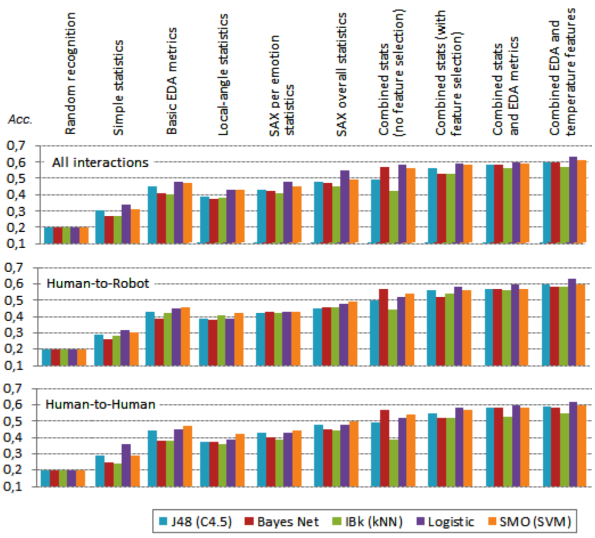

Figure 2: Emotion recognition accuracy (out of 5 emotions) using feature-driven models

stance, kNN tends to select the features from a sole scenario when $k<4$, while C4.5 trees have dedicated branches for each scenario.

Note, additionally, that these accuracy levels also reveal the adequacy of emotion description models, which can simply rely on centroid and dispersion metrics over the most discriminative features.

Additionally, to understand the relevance of features extracted from SAX and local-angle representations to differentiate emotions under assessment, one-way ANOVA tests were applied with the Tukey post-hoc analysis. A significance of 5% was considered for the Levene's test of variance homogeneity, ANOVA and Tukey tests. Both features derived from the raw, SAX and local-angle electrodermal signals were considered. A representative set of electrodermal features able to separate emotions is synthesized in Table 1.

Gradient plus centroid metrics from SAX signals can be adopted to separate negative emotions. Dispersion metrics from local-angle representations differentiate positive emotions. Rise time and response amplitude can be used to isolate specific emotions, and statistical features, such as median and distortion, to predict the affective valence. Kurtosis, which reveals the flatness of the response's major peak, and features derived from the temperature signal were also able to differentiate emotions with significance using the proposed representations.

Principles #3 and #4. In our experimental setting, the inclusion of sequential behavior leads to an increase of accuracy levels nearly 10pp. The output of HMMs were, additionally, combined the output of

Features (with strongest statistical significance to differentiate emotions' sets)	***Separated emotions***
Accentuated *dispersion* metrics (as the mean root square error) from the SAX and local-angle representations	Positive (empathy, expectation, surprise)
Median (relevant to quantify the sustenance of peaks), *distortion* and *recovery time* from SAX signals	Positive from negative from neutral emotions
Gradient (revealing long-term sympathetic activation by measuring the EDA baseline changed) and *centroid* metrics from SAX signals	Fear from frustration
Rise time	Empathy from others
Response amplitude	Surprise from others

Table 1: Features with potential to discriminate emotions

			SAX signal	Inc. local-angle	Inc. temperature	Inc. features
HMM (fully connected architecture)	Recognition accuracy	All	0.40	0.42	0.46	0.67
		Robot	0.39	0.41	0.44	0.66
		Human	0.39	0.42	0.45	0.67
	Discrimination accuracy	All	0.86	0.88	0.89	–
		Robot	0.87	0.88	0.91	–
		Human	0.86	0.88	0.90	–
HMM (left-to-right architecture) (Murphy, 2002)	Recognition accuracy	All	0.43	0.44	0.48	0.71
		Robot	0.42	0.43	0.47	0.71
		Human	0.41	0.44	0.47	0.69
	Discrimination accuracy	All	0.87	0.88	0.90	–
		Robot	0.87	0.89	0.90	–
		Human	0.87	0.88	0.89	–

Table 2: Accuracy of sequence learners to recognize an emotion (out of 5 emotions) and to correctly discard the 3 least probable emotions

probabilistic feature-based classifiers (logistic learners were the choice). Table 2 discloses the results when adopting HMMs with alternative architectures for approximately 30 signals per emotion (empathy, expectation, surprise, stress, frustration).

Interestingly, the learned HMMs are highly prone to accurately neglect 3 emotion labels that do not fit in the learned behavior. In particular, left-to-right HMM architectures are particularly well-suited to mine SAX-based signals. Note, additionally, that left-to-right architectures are a good emotion descriptor due to the high interpretability of the most probable behavior of the signal when disclosing the most probable emissions along the main path. Similar architectures can be implemented by controlling the initial transition and emission probabilities.

Although the local-angle representation is not as

Myers-Briggs type	***Correlated features*** ([+] positive correlation; [–] negative correlation)
Extrovert-introvert	[+] Dispersion metrics of SAX signal [–] Centroid metrics of SAX signal [–] Response amplitude
Sensing-intuition	[–] Dispersion metrics of raw and SAX signal [–] Dispersion metrics of local-angles [–] Rise time
Feeling-thinking	[+] Median and dispersion metrics of SAX signal [–] Declive and centroid metrics of local-angles [–] Rise time
Judging-perceiving	[–] Centroid metrics of raw signal [–] Dispersion metrics of SAX signal [+] Response amplitude

Table 3: Influence of subjects' profile on EDA responses

critical as SAX for sequential learning, its weighted use for emotion recognition and discrimination has a positive impact in the accuracy levels.

The why behind the success of adopting HMMs for emotion recognition resides on their ability to: *i)* detect flexible behavior, such as peak-sustaining values and fluctuations (hardly measured by features); *ii)* to cope with individual differences (with the SAX scaling strategy being done with respect to all stimuli, to the target stimulus, to all subjects or to subject-specific responses); *iii)* to cope with subtle variation using the local-angle representation is used as the input signal; and *iv)* to deal with lengthy responses (by increasing the number of hidden states). Additionally, HMMs can easily capture either a smoothed behavior or a more delineated behavior by controlling the signal cardinality using SAX.

Principle #5. Pearson correlations were tested to correlate the physiological expression with the subjects profile. This analysis, illustrated in Table 3, shows that their inclusion can be a critical input to guide the learning task. A positive (negative) correlation means that higher (lower) values for the assessed feature are related with a polarization towards either the extrovert, sensing, feeling or perceiving type.

We can observe, for instance, that responses from sensers and feelers are quicker, while extroverts have a more instable signal (higher dispersion) although less intense (lower amplitude).

The insertion of the relative score for the four Myers-Briggs types was found to increase the accuracy of IBk, who tend to select responses from subjects with related profile. Also, for non-lazy probabilistic learners, four data partitions were created, with the first separating extroverts from introverts and so on. One model was learned for each profile. Recognition for a test instance now relies on the equally weighted combined output of each model, which result in an increased accuracy of 2-3pp. Although the improvement seems to be subtle, note that the split of instances hampered the learning of the type-oriented models since we are relying on small-to-medium number of collected signals.

Principle #6. The analysis of the variance of key features and of the learned generative models per emotion provide critical insights for further adaptations of the learning task. For instance, the variance of rising time across subjects for positive-surprise was observed to be high due to the fact that some subjects tend to experience a short period of distrust. The inclusion of similar features in logistic model trees, where a feature can be tested multiple times using different values, revealed that they tend to be often selected, and, therefore, should not be removed due to their high variance.

Another illustrative observation was the weak convergence of the Markov model for empathy due to its idiosyncratic expression. Under this knowledge, we adapted the left-to-right architecture to include three main paths. After learning this new model, we verified a heightened convergence of the model for each one of the empathy paths, revealing three distinct forms of physiological expression and, consequently, an improved recognition rate.

Principle #7. We performed additional tests to understand the impact of the experimental conditions on the physiological expression of emotions. First, we performed a t-test to assess the influence of features derived from the signal collected during all the affective interaction (without partitions by stimulus) on the adopted type of interaction (human-to-human vs. human-to-robot). Results over the SAX representation show that human-to-human interactions (in comparison to human-to-robot) have significantly: *i)* a higher median (revealing an increased ability to sustain peaks), and *ii)* higher values of dispersion and kurtosis (revealing heightened emotional response).

Second, we studied the impact of the subjects' perception on the experiment by correlating signal features with the answers to a survey made at the end of the interaction. Bivariate Pearson correlation between a set of scored variables assessed in the final survey and physiological features was performed at a 5% significance level. Table 4 synthesizes the most significant correlations found. They include positive correlation of local-angle dispersion (revealing changes in the gradient) with intensity, felt influence and perceived intention; positive correlation of SAX dispersion (revealing heightened variations from the

Origin	***Correlations*** with higher statistical significance
Local-angle features	[+] Dispersion metrics with the felt intensity, the understanding of the agent's intention, and his level of influence on felt emotions.
SAX-based features	[+] Dispersion metrics with the perceived empathy, trust and confidence of the agent.
Computed metrics	[+/-] Amplitude positively corr. with the perceived agent influence and negatively corr. with the felt pleasure; [-] Rise time with the perceived positivism on the agent's attitude.

Table 4: Influence of subject perception in the physiological expression of emotions

baseline) with the perceived empathy, confidence and trust; quicker rise time for heightened perceived optimism; and higher amplitude of responses for heightened felt influence and low levels of pleasure.

These two observations motivate the need to turn the learning models sensitive to additional information related with experimental conditions and with the subject perception and expectations. Their inclusion as new features in feature-based learners resulted in a generalized improved accuracy (3-5pp).

5 Conclusion

This work provides seven important principles on how to recognize and describe emotions during affective interactions from physiological signals. These principles aim to overcome the limitations of existing emotion-centered methods to mine signals. We propose the use of expressive signal representations to correct individual differences and to account for subtle variations, and the integration of sequential and feature-based models. Additionally, we demonstrate the relevance of using the traits of the participant, information regarding the experimental conditions, and specific properties of the learned models to improve the learning task.

We presented initial empirical evidence that supports the utility for each one the enumerated principles. In particular, we observed that the adoption of techniques to incorporate the seven principles can improve emotion recognition rates by 20pp. Finally, a new methodology was proposed to guide the inclusion of these principles on the learning task.

Acknowledgment

This work was supported by Fundação para a Ciência e a Tecnologia under the project PEst-OE/EEI/LA0021/2013 and PhD grant SFRH/BD/75924/2011, and by the project EMOTE from the EU 7thFramework Program (FP7/2007-2013).

REFERENCES

Andreassi, J. (2007). *Psychophysiology: Human Behavior And Physiological Response*. Lawrence Erlbaum.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Inf. Science and Stat.)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18.

Jerritta, S., Murugappan, M., Nagarajan, R., and Wan, K. (2011). Physiological signals based human emotion recognition: a review. In *CSPA, 2011 IEEE 7th International Colloquium on*, pages 410 –415.

Kulic, D. and Croft, E. A. (2007). Affective state estimation for human-robot interaction. *Trans. Rob.*, 23(5):991–1000.

Leite, I., Henriques, R., Martinho, C., and Paiva, A. (2013). Sensors in the wild: Exploring electrodermal activity in child-robot interaction. In *HRI*, pages 41–48. ACM/IEEE.

Lessard, C. S. (2006). *Signal Processing of Random Physiological Signals*. S.Lectures on Biomedical Eng. Morgan and Claypool Publishers.

Lin, J., Keogh, E., Lonardi, S., and Chiu, B. (2003a). A symbolic representation of time series, with implications for streaming algorithms. In *ACM SIGMOD workshop on DMKD*, pages 2–11, NY, USA. ACM.

Lin, J., Keogh, E. J., Lonardi, S., and chi Chiu, B. Y. (2003b). A symbolic representation of time series, with implications for streaming algorithms. In Zaki, M. J. and Aggarwal, C. C., editors, *DMKD*, pages 2–11. ACM.

Murphy, K. (2002). *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, UC Berkeley, CS Division.

Picard, R. W., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(10):1175–1191.

Rani, P., Liu, C., Sarkar, N., and Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Anal. Appl.*, 9(1):58–69.

Wagner, J., Kim, J., and Andre, E. (2005). From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *ICME*, pages 940 –943. IEEE.

Wu, C.-K., Chung, P.-C., and Wang, C.-J. (2011). Extracting coherent emotion elicited segments from physiological signals. In *WACI*, pages 1–6. IEEE.