Combining Worthless Sensor Data

Cornelia Setz, Johannes Schumm, Claudia Lorenz, Bert Arnrich, Gerhard Tröster
ETH Zurich, Wearable Computing Laboratory
Gloriastrasse 35, 8092 Zurich
setz@ife.ee.ethz.ch

ABSTRACT

Previous work on emotion recognition from physiology has rarely addressed the problem of missing data. However, data loss due to artifacts is a frequent phenomenon in practical mobile applications. Discarding the whole data instance if only a part is corrupted results in a substantial loss of data. To address this problem, we investigated two methods for handling missing data: imputation and reduced-feature models using ensemble classifier systems. The five emotions amusement, anger, contentment, neutral and sadness were elicited in 20 subjects by film clips while six physiological signals (ECG, EMG, EOG, EDA, respiration and finger temperature) were recorded. Results show that classifier fusion increases the recognition accuracy in comparison to single classifiers using imputation by up to 16.3%. We were able to analyze 100% of the data even though only 47% of the data was artifact free. Since more artifacts are expected in a mobile environment than in the laboratory, the proposed methods are especially beneficial for mobile settings.

Categories and Subject Descriptors

G.4 [**MATLAB**]

General Terms

Algorithms, Measurement, Experimentation.

Keywords

Ensemble classifier systems, classifier fusion, emotion recognition, missing data, artifacts, physiology

1. INTRODUCTION AND MOTIVATION

Applications for emotion recognition are predominantly found in the field of Human-Computer Interaction (HCI). By including emotions, HCI shall become more natural, i.e. more similar to human-human interactions where information is not only transmitted by the semantic content of words but also by emotional signaling in prosody, facial expression and gesture. Several research groups have therefore employed pattern recognition methods in order to automatically distinguish between different emotional states of the user. Modalities which have been used to detect emotions include facial expression (e.g. [1]) speech (e.g. [8]) and physiological signals ([5], [6], [7], [9]).

Our work originates from the European research project SEAT which aims at increasing comfort for airplane passengers by developing a smart seat which is able to detect the emotional state of the passenger. Since the subjective appraisal of comfort is very important, emotion recognition can help in determining appropriate adjustments of the seat, the environment, the entertainment system or the attendance of the stewardess for each passenger individually. We therefore integrated physiological sensors into an airplane seat in order to infer the emotional state of the passenger.

MobileHCI'09, September 15 - 18, 2009, Bonn, Germany.
ACM 978-1-60558-281-8.

When equipping the airplane seat with sensors, we aimed at minimizing discomfort induced by the sensor attachment. These unobtrusive sensors are generally more prone to artifacts and therefore lead to a lower signal quality. This trade-off is not restricted to our airplane seat but is relevant for any system to be used in real life. When moving from the static laboratory setting to mobile real-life, a lot of "worthless sensor data" is expected. Missing values therefore represent a serious problem for practical applications. However, this problem has so far gained little attention in emotion recognition from physiology. It is current practice to discard the corresponding episodes when encountering artifacts in the data. As an example, 20% of the data in [7] was partly corrupt and had to be discarded. In our research, we are investigating methods for handling missing values to reduce the amount of data which needs to be discarded during run-time. By one of our experiments we show how all the data in a data set can be used even though only 47% of the episodes are free of artifacts.

In the following section we present common kinds of artifacts which can be detected by plausibility analyses. In section 3 we describe methods for handling this artifact affected data. An emotion recognition experiment is described in Section 4 and the results are presented and discussed in Section 5.

2. ARTIFACTS

Artifacts are a common problem encountered when investigating physiological signals. There are different kinds of artifacts. Physiological signals may be corrupted by power line interference, motion artifacts or electrode contact noise. These artifacts can be encountered in laboratory experiments, but they will become far more problematic in mobile settings.

An example of a corrupted EDA signal is shown in Figure 1. Finger movement renders the signal unusable.

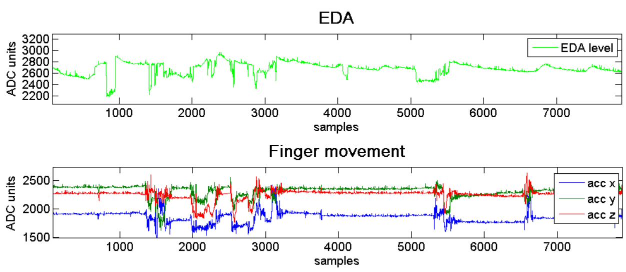

Figure 1. Artifacts in EDA signal due to finger movement.

Since we aim at recognizing emotions in everyday life, the employed sensors need to be unobtrusive. However, there is usually a trade-off between unobtrusiveness and signal quality of a sensor: the less obtrusive a sensor, the lower the expected signal quality. The RWTH Aachen has developed a novel comfortable ECG system which measures the ECG capacitively without skin contact, even through clothing [14]. Such a system was integrated in our airplane seat's backrest. Figure 2 shows a comparison between the ECG signal of the comfortable system and the normal ECG signal (measured simultaneously with wet electrodes on the chest). The comparison illustrates the trade-off between comfort and signal quality: the signal of the contactless (unobtrusive) system can not be used when the subject moves while the normal ECG is less affected.

Figure 2 also shows how plausibility analyses can determine whether the calculated R-peaks are trustworthy (marked by a green circle) or not (marked in red). By comparing each calculated RR-interval with the RR-intervals in its vicinity, implausible intervals are detected and corrupted sections are identified in a reliable way. This method is based on G.D. Clifford's work [2]; refer to [3] for Matlab Code available under the GNU public license.

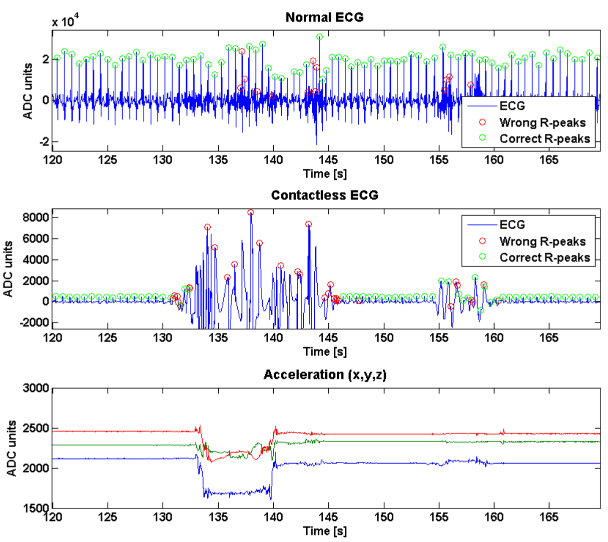

Figure 2. Movement artifacts in normal and contactless ECG.

3. METHODS

This section describes methods for handling missing values in emotion classification. We consider these methods especially promising for mobile applications.

3.1 Handling missing values

In [12], several courses of actions for handling missing data are summarized: Discard instances (in a mobile setting, this means to discard a large amount of data), acquire missing values (infeasible in automatic emotion recognition), imputation or employing reduced-feature models. The latter two are investigated in the following. Imputation uses an estimation of the missing feature or of its distribution to generate predictions from a given classifier model. A simple possibility used in practice comprises the substitution of missing values by the corresponding mean value of the uncorrupted samples.

Imputation is needed if the applied classifier employs a feature whose value is missing. The reduced-feature models technique represents an alternative approach: Instead of imputation, a new model (i.e. a new classifier) is trained which employs only the available features. A simple way to create a reduced-feature model is to use ensemble classifier systems, as described in the following subsection.

3.1.1 Ensemble classifier systems

Ensemble classifier systems consist of several classifiers whose decisions are fused to arrive at a final decision [10]. The idea behind is to consult "several experts". It can be compared to the natural behavior of humans to seek a second (or third) opinion before making an important decision. The classifier ensemble system used in this work consists of one classifier per signal modality; each signal modality thus represents an "expert". In case of encountering a missing feature in a certain signal modality, no classifier is trained for that modality and the final decision is only based on the modalities which are artifact free.

Numerous ways to combine the decisions of several classifiers exist. We have chosen and evaluated the following two simple approaches:

1. **Majority voting:** The class that receives the highest number of votes is chosen. If several classes receive an equal number of votes, the corresponding classifiers are identified as candidate classifiers and the classifier with the highest confidence among them provides the deciding vote.

2. **Confidence voting:** The class of the classifier with the highest confidence is chosen.

For the two described classifier fusion schemes, a suitable underlying classifier is needed. Linear and Quadratic Discriminant Analysis (LDA and QDA) with diagonal covariance matrix estimation were chosen since they need only few parameters to be estimated. The estimated posterior probability for the selected class was taken as confidence measure for the classifier fusion.

4. EXPERIMENTS

This section describes one of our emotion recognition experiments to show the performance of the discussed methods for handling missing data. The full experiments and results have been accepted for publication in [13].

The emotions to be recognized were chosen according to the well-known 2-dimensional emotion model of arousal and valence often used in emotion recognition studies [4]. One emotion in each quadrant plus neutral were selected as shown in Figure 3: amusement (high arousal, pos. valence), anger (high arousal, neg. valence), contentment (low arousal, pos. valence), sadness (low arousal, neg. valence) and neutral (medium arousal, zero valence).

Film clips were chosen for emotion elicitation because they are capable of eliciting strong emotional responses under standardized conditions and generate a rather long stimulation (1-10 minutes) [11].

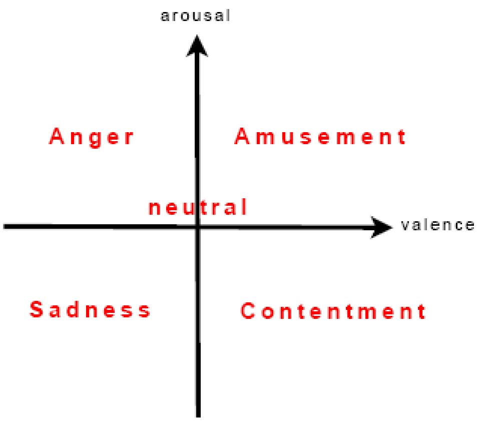

Figure 3. Emotions to be recognized in arousal valence space.

For each of the five emotions, one film clip was selected as shown in Table 1. The used film clips are either suggested in [11] (marked with * in Table 1), proposed in another paper (John Q from [6]) or self chosen.

Table 1. Film sets used in the study: Films marked with * originate from [11]. John Q was suggested in [6].

Emotion	Film Clip
Sadness	John Q
Amusement	When Harry met Sally*
Neutral	Fireplace
Anger	Cry Freedom*
Contentment	Winged Migration

Twenty (12 male, 8 female) participants were recruited for the study. The five film clips were presented to the subjects by video glasses (see Figure 4) which were not taken off during the whole experiment. After each film clip, a three minute recovery period followed. The screen went black and the subjects were requested to clean their mind and calm down.

Figure 4. Participant wearing video glasses

4.1 Signals and Features

The following physiological signals have been recorded: The Electrocardiogram (ECG), the vertical component of the Electrooculogram (EOG) to detect the blinking rate, the Electromyogram (EMG) of the Musculus Zygomaticus (muscle between mouth and eye, contracted when smiling), the tonic and phasic part of the Electrodermal activity (EDA), the respiration and the finger temperature.

Respiration and EDA were recorded with custom electronics integrated into the airplane seat. The respiration was measured by a stretch sensor attached to the seat belt (see Figure 5). The electronic circuits for the EDA directly filter and split the signal into a tonic and a phasic part.

Figure 5. EDA measurement at the left hand, respiration sensor in the seat belt. The additional electrode at the right hand can be applied to measure ECG. It was not used due to technical problems encountered shortly before experiment begin.

The ECG, the vertical EOG, the EMG of the Musculus zygomaticus and the finger temperature were measured by the commercial Mobi device of TMS International [15].

A total of 53 features was calculated from the six signals. The features were computed for the periods of the film clips (emotion phases) and for the preceding recovery periods. The features calculated during the recovery periods served as baseline values for calculating relative features by division or subtraction.

The recorded data set contained a rather large set of corrupted segments, due to technical problems and artifacts. The artifacts included: (1) EDA signal reached saturation of the amplifier during the course of the experiment; (2) R-peaks in ECG signal were erroneously detected because of a high T-wave or due to motion artifacts; (3) blinking was not visible in EOG signal due to dry skin. The problematic data sets were identified and declared as invalid. This resulted in 96% correct data for EOG, 78% for EDA, 64% for ECG and 100% for the remaining modalities. The percentage of data containing valid features of all modalities amounted to 47%.

5. Results and Discussion

In this section we present how imputation techniques and classifier fusion methods handle the case of missing feature values. The classifiers presented in the following were evaluated in a leave-one-person-out cross-validation and all the six signal modalities were used for this analysis.

Table 2. Comparison of ensemble classifier systems and single classifiers (with imputation) for LDA and QDA. The results are presented for 5 and for 4 classes (neutral excluded). Classifier fusion yields considerable benefits.

Fusion method	Classi-fier	Imputation Yes/No	Acc. 5 classes	Acc. 4 classes
None	LDA	Yes	45.0%	52.5%
Maj. Voting	LDA	No	41.0%	55.0%
Confidence	LDA	No	**47%**	**58.8%**
None	QDA	Yes	35%	40%
Maj. Voting	QDA	No	**49%**	**56.3%**
Confidence	QDA	No	48.0%	**56.3%**

The results for all 5 classes and for 4 classes (all except neutral) are presented in Table 2. Clearly, classifier fusion always yields a considerable benefit in comparison to single classifiers using all the features (with imputation). In the best case (4 classes, QDA) the increase in accuracy amounts to 16.3%.

Another interesting observation is that confidence voting always performs better then majority voting for LDA, while the tendency is reversed for QDA. The best classifier fusion technique thus seems to depend on the underlying classifiers.

A problem of our majority voting scheme was identified during this analysis: Since the majority voting is initially performed without considering the confidence values, it can happen, that several “weak” classifiers agree on a wrong decision and thereby dominate another classifier which might exhibit a high confidence value. A possibility to circumvent this problem would be to use the confidence values as weights during the majority voting process already.

5.1 Conclusion

In practical applications, data loss due to artifacts occurs frequently. Many of these artifacts can be detected automatically by plausibility analyses (e.g. unrealistic RR-intervals). Often the artifacts do not occur in all physiological signals simultaneously and discarding all the data instances containing invalid features results in a substantial amount of unusable data. In our experiment, more than half of the data would have been lost if no strategy to handle missing values had been employed. With the proposed methods we were able to analyze 100% of the data of the laboratory experiment. In a mobile environment, we anticipate even more artifacts than in the laboratory. We therefore expect our methods to be beneficial especially for mobile settings.

Two methods for handling missing features in combination with two classifier fusion approaches have been investigated. Classifier fusion has been shown to significantly increase the recognition accuracies. A maximum increase in accuracy of 16.3% was observed when comparing an ensemble classifier system to a single classifier using imputation. Whether majority or confidence voting performs better depends on the underlying classifier.

6. ACKNOWLEDGMENTS

This project is sponsored by the European Commission DG H.3 Research, Aeronautics Unit under the 6th Framework Programme under contract Number: AST5-CT-2006-030958.

7. REFERENCES

[1] BUSSO, C., DENG, Z., YILDIRIM, S., BULUT, M., MIN LEE, C., KAZEMZADEH, A., LEE, S., NEUMANN, U., NARAYANAN S. 2004. Analysis of emotion recognition using facial expressions, speech and multimodal information. In ICMI ’04: Proceedings of the 6th international conference on Multimodal interfaces. 205–211.

[2] CLIFFORD, G.D., MCSHARRY P.E., TARASSENKO L. 2002. Characterizing artefact in the normal human 24-hour RR time series to aid identification and artificial replication of circadian variations in human beat to beat heart rate using a simple threshold. Computers in Cardiology, 129-132.

[3] CLIFFORD, G.D. http://www.mit.edu/~gari/code.html.

[4] COWIE, R., DOUGLAS-COWIE, E., TSAPATSOULIS, N., VOTSIS, G., KOLLIAS, S., FELLENZ, W., AND TAYLOR, J. 2001. Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine 18, 1 (January), 32–80.

[5] KIM, J. AND ANDRÉ, E. 2008. Emotion recognition based on physiological changes in music listening. IEEE Trans. Pattern Anal. Mach. Intell. 30, 12, 2067–2083.

[6] KREIBIG, S. D., WILHELM, F. H., ROTH, W. T., GROSS, J. J. 2007. Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. Psychophysiology 44, 787–806.

[7] LISETTI, C. L., NASOZ, F. 2004. Using noninvasive wearable computers to recognize human emotions from physiological signals. EURASIP J. Appl. Signal Process. 2004, 1, 1672–1687.

[8] NEIBERG, D., ELENIUS, K., LASKOWSKI, K. 2006. Emotion recognition in spontaneous speech using gmms. In ICSLP ’06: 9th International Conference on Spoken Language Processing. 809–812.

[9] PICARD, R. W., VYZAS, E., HEALEY, J. 2001. Toward machine emotional intelligence: Analysis of affective physiological state. IEEE Trans. Pattern Anal. Mach. Intell. 23, 10, 1175–1191.

[10] POLIKAR, R. 2006. Ensemble based systems in decision making. IEEE Circuits and Systems Magazine 6, 3, 21–45.

[11] ROTTENBERG, J., RAY, R. R., GROSS, J. J. 2007. Emotion elicitation using films. In J. A. Coan and J. J. B Allen (Eds.), The handbook of emotion elicitation and assessment. New York: Oxford University Press, 9–28.

[12] SAAR-TSECHANSKY, M. AND PROVOST, F. 2007. Handling missing values when applying classification models. J. Mach. Learn. Res. 8, 1623–1657.

[13] SETZ, C., SCHUMM, J., LORENZ, C., ARNRICH, B., TRÖSTER, G. 2009. Using Ensemble Classifier Systems for Handling Missing Data in Emotion Recognition from Physiology: One Step Towards a Practical System. Accepted for the 2009 International Conference on Affective Computing & Intelligent Interaction.

[14] STEFFEN, M., ALEKSANDROWICZ, A., LEONHARDT, S. Mobile Noncontact Monitoring of Heart and Lung Activity. 2007. IEEE Trans. Biomed. Circuits Syst. 1, 4, 250–257.

[15] TMS International. http://www.tmsi.com.