Generic Features for Biosignal Classification

Ulf Jensen, Matthias Ring, Bjoern Eskofier

Digital Sports Group, Pattern Recognition Lab, University Erlangen-Nuremberg, Germany

Introduction

The recent progress in sensor technology in terms of better signal acquisition, lower energy consumption and higher integration ability paves the way for a variety of mobile data collection and analysis applications. From a sports perspective, this enables wearable support and monitoring tools that are often realized as Body Sensor Networks (see fig. 1). Different biosignals, like physiological and kinematic data, can be acquired with such networks and pattern recognition methods provide valuable tools for online and offline signal analysis (Eskofier et al., 2009).

Figure 1: Example of a Body Sensor Network as athlete support system.

However, defining a suitable representation of the measured data often requires expertise in signal processing or complex transformations. In contrast to image analysis, standard feature sets for low-dimensional biosignals have not been established. In addition, some proposed transformations are computationally complex and therefore not suitable for mobile applications (Ciaccio el al., 1993). In a mobile embedded context, resources are limited and algorithm development is constrained by memory, computing resources and runtime.
Therefore, we compiled a set of simple, generic features and applied it to two classification problems in the field of biosignal analysis. First, we classified physical activity levels using physiological electrocardiogram (ECG) data. Activity levels are important to assess the demands of a sport or to support an athlete in remaining in a desired training zone. Second, we used inertial golf putt data to classify a player's experience level. Such a system can be used to recommend equipment, assign suitable training courses or detect specific error patterns. The golf data were of

event-based nature as we extracted single putts. The ECG data, in contrast, consisted of continuous measurements for a longer period of time.
The developed feature set was integrated in a freely available software package that supports the development of embedded classification systems (Ring et al., 2012). Therefore, the software facilitates a rapid prototyping of classification systems from a sports expert perspective.

Methods

Hardware

We used the SHIMMER™ sensor platform for data collection (McGrath et al., 2009). The sensor nodes were equipped with an ECG sensor module for the collection of physiological data and equipped with a gyroscope sensor module for kinematic data collection. Data were wirelessly transmitted to a laptop for recording and analysis.

Data

We used two datasets to show the applicability of the proposed feature set. First, ECG data were classified for the purpose of activity recognition. Second, kinematic sensor data collected during putting were analyzed for experience level classification.
The ECG data consisted of a 4-lead measurement with 100 Hz sampling rate. Data during two activity states, sitting and walking, were collected. Five subjects volunteered in the study and were recorded for two minutes during each activity. The measured signal was cut in intervals of 200 samples each and intentionally not preprocessed in any way. This results in 60 patterns per activity and person, thus, a total sum of 600 patterns. Features were computed on each of these intervals.
The kinematic data consisted of inertial sensor data collected at the golf club head during putting and contained 3-D accelerometer and 3-D gyroscope data. Eleven either experienced or completely inexperienced subjects participated in the study. Overall, 315 putts from various distances and with different equipment were collected. We developed an automatic detection and segmentation algorithm to extract the putts (Jensen et al., 2011) and computed features for each of them. This resulted in 315 patterns overall. The data were previously classified with a different feature set. These features, in contrast to the presented set, required prior knowledge and were defined by golf experts (Jensen et al., 2012).

Features

Our feature set consisted of statistical moments and additional simple signal characteristics. We intentionally used solely measurement and structure features to avoid complex computations. The nine features are displayed in tab. 1.

Each pattern of the physiological data had a dimension of nine because the analysis was based on a 1-D measurement. In contrast, the patterns of the kinematic data analysis problem had a dimensionality of 54 as nine features were drawn from each of the six measurement dimensions.

Table 1: Compiled set of nine generic features consisting of statistical moments and signal characteristics.

Statistical moments	Signal characteristics
1. Mean	6. Minimum
2. Standard deviation	7. Maximum
3. Variance	8. Energy
4. Kurtosis	9. Median
5. Skewness	

Analysis

We compared a range of classifiers in the analysis as no classifier is known to outperform others in any classification problem (Theodoridis et al., 2009). See tab. 2 for a complete list of classifiers and a short description. Detailed descriptions of the classifiers and the processing pipeline of classification systems can be found in (Jain et al., 2000) and (Theodoridis et al., 2009).

Table 2: Overview and description of applied classifiers.

Name	Description
AdaBoost (AB)	AB iteratively combines several weak classifiers to a non-linear decision boundary.
Linear Discriminant Analysis (LDA)	LDA constructs a decision boundary by minimizing the intra-class and maximizing the inter-class variability.
Naive Bayes (NB)	According to the Bayes rule, patterns are assigned to the class with the highest posterior probability.
Nearest Neighbor (NN)	Patterns are assigned to the class of the nearest (Euclidean distance) training pattern.
Support Vector Machine, linear kernel (SVM)	SVM constructs a decision boundary with a maximal margin to separate different classes.

The datasets were analyzed in a leave-one-subject-out procedure to estimate the classification performance. In this procedure, a classifier was trained on data from all subjects except one and this one subject was used as test set. Thus, data from one subject was either used in classifier training or testing but never in both steps. This procedure was repeated for all subjects and ensured that the classifier did not learn individual subject properties. The resulting subject-dependant classification rates were averaged to an overall classification rate.
We used custom software based on the WEKA software package (Hall et al., 2009) for classification. Our software (Ring et al., 2012), the Embedded Classification Software Toolbox (ECST), enhances the WEKA software by a classifier cost analy-

sis and also includes the extraction of generic features as described in this article. See fig. 2 for a screenshot of the ECST.

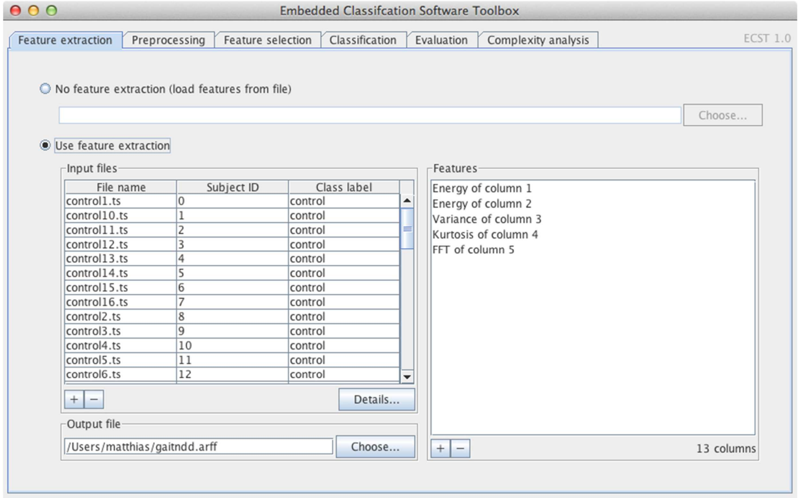

Figure 2: Screenshot of the feature extraction step of the Embedded Classification Software Toolbox.

Experiments

Using the ECG data and the proposed feature set, we built a system to classify two activity states with a granularity of two seconds.
The golf putt dataset was used to classify experienced and unexperienced players. First, this was performed with the generic feature set. Second, to compare the results, the experiment was conducted with a golf-specific feature set (Jensen et al., 2012) that is compiled in tab. 3.

Table 3: Specific feature set for golf experience classification based on different putt phases like back swing (BS), forward swing (FS) and follow-through (FT).

Number	Description
1-4,7	Duration of {BS, FS, BS&FS, FT, complete putt}
5-6	Temporal ratio of {FS&FT, BS&FS}
8-10	Rotation in BS and FS (x-, y-, z-axis)
11-13	Rotation in pre impact phase (x-, y-, z-axis)
14-16	Rotation in post impact phase (x-, y-, z-axis)
17-19	Rotation angle in {BS, FS, FT}
20	Rotation angle ratio of FS and FT
21-22	Acceleration change in {pre, post} impact phase
23-24	Velocity on impact (rotation, horizontal)
25-26	Acceleration maximum in FS (position, value)
27-28	Velocity maximum in FS (position, value)

Results

Tab. 4 gives an overview of the results for the activity classification using ECG data. The LDA classifier performed best and all classifiers were able to reach a classification rate of 80% or more.
The results from the golf experience classification are compiled in tab. 5. In the classification with generic features, LDA outperforms the other classifiers. If compared to the expert feature set, LDA and NN showed better results while the classification rates of the remaining classifiers were lower. The best overall result reached the LDA classifier with the generic feature set.

Table 4: Classification rates (in percent) of the activity classification with ECG data.

Classifier	AB	LDA	NB	NN	SVM
Generic Features	84.8	88.8	80.5	83.7	81.2

Table 5: Classification rates (in percent) of the golf experience classification with inertial sensor data.

Classifier	AB	LDA	NB	NN	SVM
Generic Features	69.9	90.2	70.2	74.4	75.8
Expert Features	76.7	82.4	86.1	68.4	78.8

Discussion

We were able to solve two classification problems in the field of biosignal analysis with a simple and generic feature set. Two different types of signals, a continuous 1-D ECG signal and an event-based 6-D kinematic signal were considered. In both datasets, the LDA classifier performed best. Thus, the characteristic of the generic feature set seems to fit well for this classifier.
In the activity classification problem, all classifiers were able to successfully classify over 80% of two-second-intervals to the “walking” or “sitting” class. For an activity monitoring application, the results for consecutive interval classifications can be combined to a more robust classification. The chosen interval length reflects the time resolution of the classification and needs to be adapted to the corresponding data and the desired application.
As mentioned above, the ECG signal was not preprocessed in any way. Movement and breathing artifacts as well as noise were still present in the data and the two-second processing interval was not aligned to heartbeat events. We chose this set-up intentionally to exploit the information content of the unprocessed signal. Furthermore, preprocessing requires computational effort and, therefore, demands resources on an embedded system. The dataset we presented was rather small but certainly shows the potential of the proposed generic feature set for ECG classification.

In the golf putt data classification, the LDA classifier performed best and reached a classification rate of 90.2%. Compared to the ECG classification, the difference to the other classifiers increased. It can be speculated that the performance of this classifier increases with the number of sensors and, therefore, the dimensionality of the classification problem.

The fact that the best overall result was achieved with the generic feature set is remarkable. The results furthermore revealed that some classifiers perform better with the generic feature set while the performance of others degraded. We speculate that this was due to the statistical nature of the generic feature set. The pattern distribution might fit better or worse to the respective classification paradigms. However, the expert feature set has a major advantage over the generic feature set. Golf experts can directly interpret the feature values and a feature selection can be applied to find discriminant features for further interpretation. The generic features, in contrast, are more abstract and harder to interpret.

Summary and outlook

We presented a generic feature set consisting of mainly statistical measures that can be applied for biosignal classification in an embedded context. We proved the applicability with the successful classification of two different feature sets and showed that the generic feature set outperformed expert features in some cases.

Currently, our research focuses on the cost estimation of these features. We will integrate the memory demand and computational effort of the generic feature extraction into the ECST to be able to benchmark a complete classification system.

Acknowledgements

The authors would like to thank Stefan Geißelsöder for data collection. This work was funded by the Bavarian Ministry for Economic Affairs, Infrastructure, Transport and Technology and the European Fund for Regional Development.

References

Eskofier, B., Oleson, M., DiBenedetto, C. & Hornegger, J. (2009). Embedded Surface Classification in Digital Sports. Pattern Recognition Letters, 30 (16), 1448–1456.

Ciaccio, E., Dunn, S., & Akay, M. (1993). Biosignal Pattern Recognition and Interpretation Systems. IEEE Engineering in Medicine and Biology, 12 (4), 106–113.

Ring, M., Jensen, U., Kugler, P. & Eskofier, B. (2012). Software-based Complexity Analysis for the Design of Embedded Classification Systems. In Proc. of the ICPR 2012 (accepted for publication), Tsukuba.

McGrath, M. & Dishongh, T. (2009). A Common Personal Health Research Platform — SHIMMER™ and BioMOBIUS™. Intel Technology Journal, 13 (3), 122–147.

Jensen, U., Kugler, P., Dassler, F. & Eskofier, B. (2011). Sensor-based Instant Golf Putt Feedback. In Proc. of the IACSS 2011, Shanghai, 49–53.

Jensen, U., Dassler, F. & Eskofier, B. (2012). Classification of Kinematic Golf Putt Data with Emphasis on feature selection. In Proc. of the ICPR 2012 (accepted for publication), Tsukuba.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. & Witten, I. (2009). The WEKA Data Mining Software: An Update. SIGKDD Explorations, 11 (1), 10–18.
Theodoridis, S. & Koutroumbas, K. (2009). Pattern Recognition. Academic Press.
Jain, A., Duin, R. & Mao, J. (2000). Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (1), 4–37.