34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Applications of Supervised Learning to Biological Signals: ECG Signal Quality and Systemic Vascular Resistance

Stephen J. Redmond, *Member*, *IEEE*, Qim Yi Lee, *Student Member, IEEE,* Yang Xie, *Student Member, IEEE,* Nigel H. Lovell, *Fellow*, *IEEE*

***Abstract*—Discovering information encoded in non-invasively recorded biosignals which belies an individual's well-being can help facilitate the development of low-cost unobtrusive medical device technologies, or enable the unsupervised performance of physiological assessments without excessive oversight from trained clinical personnel. Although the unobtrusive or unsupervised nature of such technologies often results in less accurate measures than their invasive or supervised counterparts, this disadvantage is typically outweighed by the ability to monitor larger populations than ever before. The expected consequential benefit will be an improvement in healthcare provision and health outcomes for all. The process of discovering indicators of health in unsupervised or unobtrusive biosignal recordings, or automatically ensuring the validity and quality of such signals, is best realized when following a proven systematic methodology. This paper provides a brief tutorial review of supervised learning, which is a sub-discipline of machine learning, and discusses its application in the development of algorithms to interpret biosignals acquired in unsupervised or semi-supervised environments, with the aim of estimating well-being. Some specific examples in the disparate application areas of telehealth electrocardiogram recording and calculating post-operative systemic vascular resistance are discussed in the context of this systematic approach for information discovery.**

I. Introduction

A. Automation

Engineers strive to automate many labor intensive processes, which are usually the reserve of trained human specialists; this has certainly been the case for biomedical engineers. While certain branches of biomedical engineering continue to make advancements in sensing and transduction technologies for physiological measurement, on a parallel track is an effort to automate, standardize and even improve on human interpretation of these measurements.

B. Established clinical applications

Until recently, this effort to mimic, support and improve the human interpretation of physiological measurements has been somewhat confined to the clinical sphere. A small selection of examples include: the tedious task of interpreting electrocardiogram (ECG) recordings [1]; or classifying stages of sleep from an eight hour overnight polysomnography [2]. Computer-based interpretation of measurements is well established in many clinical areas as a support tool.

S. J. Redmond (email: s.redmond@unsw.edu.au), Y. Xie and N. H. Lovell (email: n.lovell@unsw.edu.au) are with the Graduate School of Biomedical Engineering, and Q. Y. Lee is with the School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney, NSW 2052, Australia.

C. Healthcare at a distance

A more recent research trend has pondered the usefulness of communications technologies, primarily the internet, to improve health and well-being. This mode of healthcare delivery has been termed 'telehealth'. Computer-aided interpretation of remotely acquired biosignal recordings poses a unique set of challenges related to the design of the measurement hardware and the associated interpretive software employed; some of which are discussed later.

D. Point-of-care

Another recent driver of medical device innovation has been the continuing miniaturization of transduction and processing hardware, resulting in a wide variety of point-of-care testing technologies; where the monitoring technology is easily brought to the patient. With this style of monitoring, it is desirable to deliver real-time interpretation. While many of these devices record the presence or concentration of pathogens, biomarkers or other agents in bodily fluids, point-of-care biosignal recording, such as bedside 12-lead ECG, is increasingly common.

E. Emergency care

Point-of-care diagnostics and therapeutics can also extend to community care. They are frequently used by paramedics or emergency departments to measure ECG, blood pressure (BP) and pulse oximetry (using photoplethymography (PPG). Automatic external defribrillators are becoming commonplace in public spaces. As this trend towards miniaturization advances, other tests which may currently require unwieldy apparatus will become portable.

This shift towards the miniaturization of established monitoring equipment is inspiring the development of new screening tools, which utilize these same recording modalities by uncovering hitherto unknown correlates of serious conditions, such as sepsis or stroke, encoded in these commonly acquired biosignals [3, 4].

F. Ambulatory monitoring

Yet another variation on this theme is what has become known as ambulatory monitoring. This is not a new concept, with the development of the Holter ECG monitor in the 1960s being one of the first embodiments of an ambulatory monitor [5]. While ambulatory bioelectric signal recording has progressed to the point that exercise heart rate monitors are commercially available [6], another field of ambulatory monitoring has arisen to assesses movement and mobility using miniaturized inertial sensors [7]. Current application areas attracting the greatest research attention include

activity monitoring, energy expenditure estimation, fall detection and fall risk estimation [7].

G. Efficacy versus reach and speed

While the developments and trends described above are certainly exciting, these benefits afforded by device miniaturization and ubiquitous communication are not without their disadvantages.

The principal benefits of these systems relate to their reach into the community or the improved promptness with which monitoring is performed. For example, screening of large populations for their risk of falling [8], or a very early assessment of whether someone is suffering from stroke [4], should be viewed as an improvement in healthcare.

The disadvantages alluded to relate to the inherent loss of accuracy of such systems. This usually stems from three sources: use of compromised availability of biosignals, environmental noise and movement.

For example, sleep apnea screening has moved from the sleep clinical into the home. This has greatly increased the number of people who can be screened. However, a full polysomnograph is not feasible in the home. These at-home systems compromise their accuracy to improve usability, using only a subset of required signals; for example, nasal airflow, chest movement, ECG or pulse oximetry [9, 10]. Furthermore, subject movement will greatly disrupt or destroy the recorded signals; the latter due to loss of electrode or transducer placement. When designed with high sensitivity and low specificity, these lower accuracy systems can still prove useful.

H. Commonalities

The common characteristic underpinning all of the abovementioned monitoring systems are: (i) the measurement is performed in a noisy or unsupervised environment; (ii) use a subset of the signals normally available in a clinical environment to optimally extract correlates of a desired, but unknown physiological parameter. The remainder of this paper discusses the methodology of supervised learning and gives some examples of its application in constructing an algorithm to perform automated signal interpretation.

II. A Brief Review of Supervised Learning

The following sections momentarily digress away from the discussion of medical devices to discuss the background of supervised learning [11].

A. The concept of the feature space and pattern labels

'Feature' is the term given to a characteristic of a pattern. Here, by pattern, we may mean a biosignal recorded from a human. We can extract many features from the pattern to summarize its properties. This group of extracted features is called a feature vector.

The types of features we extract should be intimately related to the recognition task. They should ideally remain uninfluenced by all other signal variations; for example, recording from two different people should ideally not influence a feature's discriminatory power due to differences between the people.

For brevity, let us consider only real-valued features. If there are d features, the d-dimensional feature vector can be visualized as a point in a d-dimensional feature space. This feature vector also has an associated label, which may the health parameter to be estimated; the label can be real, discrete or nominal valued. For example, systolic and diastolic BP are features extracted from cuff pressure and auscultatory waveforms, while the associated label might take the nominal value of 'high BP or 'low BP'.

B. Feature extraction

The method by which features are extracted is by far the most important part of the entire recognition process. The number of transformations which could be performed are infinite, but there are some broad categories of methods which can be used, including Fourier analysis, wavelet analysis, empirical mode decomposition, and morphological analysis to name but a very few. (The application of these methods in data and image compression algorithms is not coincidental, as feature extraction also aims to compress the pattern, but according to different criteria).

1) Windowing

It may be the case that the target labels change over time, in which case a sliding window is often moved across the signal and a feature vector extracted for each window placement. This leads to the notion of preprocessing.

2) Preprocessing

Preprocessing is anything which prepares the signal for feature extraction. There is no clear distinction between what is considered feature extraction and what is considered preprocessing. However, when windowing (above) is used, preprocessing can be considered as any operation applied to the signal at a global level before windowing is performed and features are extracted. Filtering is a perfect example of a preprocessing operation.

C. Classification or regression

Whether the label applied to the feature vector is real, discrete or nominal, determines the type of estimation model used. If nominal or discrete, a pattern classifier is used [11]. If real or discrete, a regression model is used; although regressing to a discrete (but possibly real-valued) feature may not work well in practice, and a classifier would be preferred.

1) Supervised classification

A classifier attempts to carve the feature space into regions (not necessarily contiguous) which group all feature vectors with the same label together. For supervised training of classifier models, a large pool of training data is required to discover these regions. The choice of mathematical function to parameterize these regions is primarily what delineates the many flavors of classifier model described in the literature. At heart they all perform the same task, with their own application-dependent advantages and disadvantages; namely, computational complexity, memory requirements and training data requirements relative to feature space dimension.

2) Regression

Linear multiple regression attempts to find a vector (the normal to a hyperplane) pointing in some direction in the feature space, such that when all the feature vectors are projected onto this vector they are arranged in approximately the same order as their real-valued labels. Phrased another way, it finds the optimal set of weights, and an additive constant, such that a weighted sum of the features gives the least-mean-square error to the real-valued target labels .

D. Model validation with finite amounts of training data

Training and testing of the model should never be performed with the same data, or the results will be optimistically biased. However, often there are only a finite number of feature vectors available to train the classifier. One might hold back some data for testing – but which?

To solve this problem, cross validation is used. Some fraction of the data, either chosen systematically or randomly, is removed and the classifier trained with the remainder. The withheld data is later used for testing. This is repeated many times, withholding different data subset each time for testing, with the averaged results used as the final performance metric.

E. Feature selection

With infinite amounts of training data, all imaginable features can be included in the feature vector, and if they do not contribute information they will be ignored. With smaller training sets, the generalized classifier performance is intimately related to the ratio of the dimension of the feature space to the number of training vectors.

Feature selection involves searching through as many feature subsets as possible to see which yield the best classifier generalization, as determined by cross validation. Other more general methods project the feature space to a lower dimensional subspace, while preserving the discriminatory information of the features.

However, care must once again be taken when performing feature selection (dimension reduction) with small training sets. Here a second outer cross validation loop is required to ensure that the features (subspace) selected are not simply optimized for the small training set available.

III. Some Illustrative Examples

The following two sections provide some specific illustrative examples of the use of supervised learning to develop automated algorithms for the interpretation of non-invasively acquired biosignals, recorded (or destined to be recorded) in unsupervised or semi-supervised environments.

A. ECG signal quality

ECG recordings acquired in unsupervised telehealth environments suffer from detrimental movement artifact and electrode contact issues. Our research group has been working to develop algorithms to detect movement artifact [12] and interpret the general quality of single-lead ECG recordings [13].

Fig. 1 shows a raw single-lead ECG signal acquired unsupervised by an elderly telehealth patient using a lead-I configuration, with the hands placed on dry electrode plates.

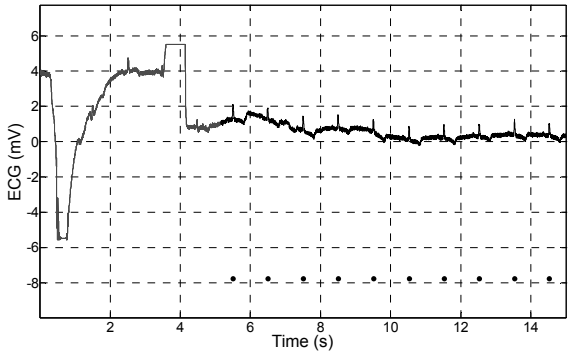

Fig. 1. A raw lead-I ECG signal acquired from an elderly telehealth patient in their own home, placing their hands on dry electrode plate electrodes. Artifact before 5 s is detected using the heuristic algorithm in [12], the quality of the remaining signal is assessed using a supervised classifier model [13].

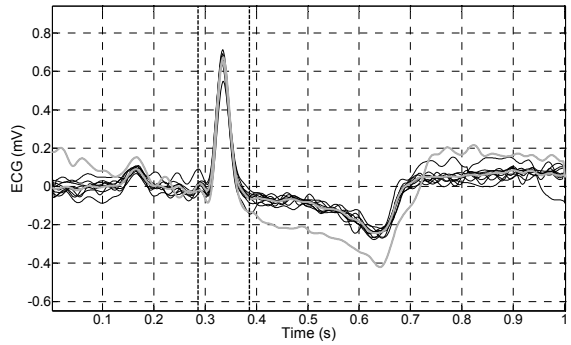

Fig. 2. ECG signal from Fig. 1 after pre-processing: filtering, QRS detection, segmentation, alignment and clustering (to account for ectopic beats). Seven features (which constitute a feature vector) are extracted from this pre-processed data for the purposes or classifying into one of three possible quality classes; 'Good', 'Average' or 'Bad'.

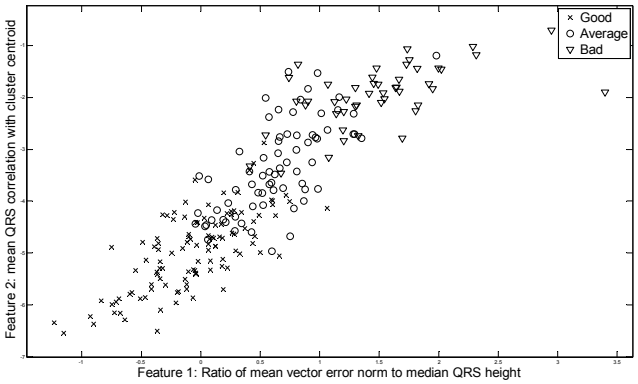

Fig. 3. Scatter plot of two of the seven extracted features (see Fig. 2) for all 300 training signals. The objective of the classifier is to segment the feature space into regions so that a new unseen feature vector (from a new ECG signal) can be classified into a quality class. If a continuous scale of quality is required, it would be more advisable to use a multiple regression model than a discrete output classifier model.

Once the artifact in the early part of the signal is removed, the remainder of the signal is pre-processed (filtering, QRS detection, segmentation, alignment and clustering) to generate Fig. 2, ready for feature extraction.

A number of features are extracted which quantify the variation between beats and the amount of noise present. Two of these features extracted from 300 such ECG signals, with qualities labels of 'Good', 'Average' and 'Bad', are plotted in Fig. 3. A supervised classifier model can then

mathematically generate a near-optimal decision rule to segment this feature space into regions according to the quality labels assigned to the training data.

B. Systemic vascular resistance

The same general methodology has been applied to non-invasively estimate systemic vascular resistance (SVR) from a PPG signal [14]. Knowledge of SVR and BP can be used to estimate cardiac output; which normally requires an invasive indicator dilution assessment or a skilled ultrasound technician.

Features are extracted from the waveform and spectrum of a 4-minute segment of the PPG. Fig. 4 shows PPG signals acquired from subjects with low and high SVR. Pre-processing is also performed to detect the beat locations, and some features are extracted from individual beats. Three of these features plotted as points in the feature space for all 48 subjects can be seen in the 3-D plot of Fig. 5. Again, a supervised pattern classifier is trained and validated using cross validation, and a feature selection search performed to highlight those features with the most discriminatory power.

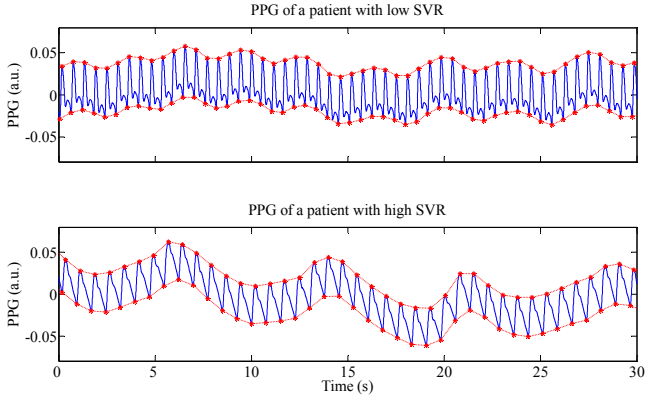

Fig. 4. Sample PPG plots for low (top) and high (bottom) SVR subjects. Generating some features involves calculating the spectrum of the signal, extracting the resulting power in different frequency bands to estimate larger low frequency variation seen in subjects with high SVR [14].

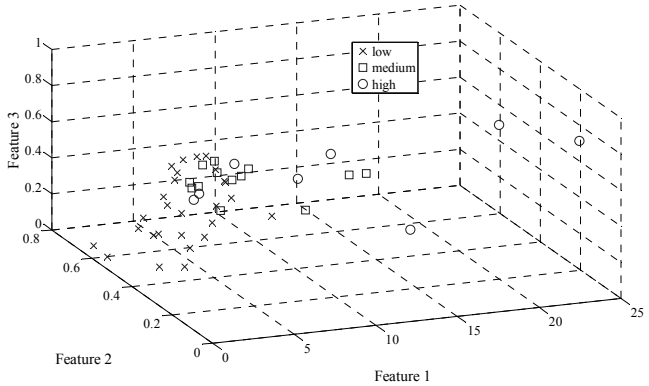

Fig. 5. Three features extracted from 48 PPG signals. Labels represent 'low', 'medium' and 'high' SVR classes. Features: 1) low-to-high frequency power ratio, 2) peak-to-notch time, 3) reflection index. [14].

IV. Discussion and Conclusion

This paper has presented a discussion on a continuing research trend which attempts to make medical device technology smaller, more portable and more intelligent. While such advancements are driven partly by progress in hardware and transduction, the unreliability of the signals acquired with such devices when used in unsupervised environments must be offset with intelligent signal processing to validate the quality of such signals.

Furthermore, only the most robust of signals may be reliably acquired in unsupervised environments. This has led a rise in research activity aiming to replicate established supervised clinical assessments using only a subset of the original signals required, or using a secondary set of signals which are indirectly influenced by the physiological system under investigation, but are easier to obtain reliably.

Finally, a tutorial overview is provided on how supervised classification methodologies may be applied to this task of discovering and automatically classifying health-related information from non-invasively acquired biosignals. Some illustrative examples in ECG signal quality validation and non-invasively estimating SVR have been provided.

V. References

[1] N. V. Thakor, Y. S. Zhu, and K. Y. Pan, "Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm," *IEEE Transactions on Biomedical Engineering,* vol. 37, pp. 837-843, 1990.

[2] R. Agarwal and J. Gotman, "Computer-assisted sleep staging," *IEEE Transactions on Biomedical Engineering,* vol. 48, pp. 1412-1423, 2001.

[3] P. M. Middleton, C. H. Tang, G. S. Chan, S. Bishop, A. V. Savkin, and N. H. Lovell, "Peripheral photoplethysmography variability analysis of sepsis patients," *Med Biol Eng Comput,* vol. 49, pp. 337-47, Mar 2011.

[4] M. F. L. Ramirez, R. T. Tibayan, C. E. Marinas, M. E. Yamamoto, and E. V. S. Caguioa, "Prognostic value of hemodynamic findings from impedance cardiography in hypertensive stroke," *Am J Hypertens,* vol. 18, pp. 65S-72S, Feb 2005.

[5] J. Gawłowska and J. K. Wranicz, "Norman J. "Jeff" Holter (1914–1983)," *Cardiology Journal,* vol. 16, pp. 1-2, 2009.

[6] R. M. T. Laukkanen and P. K. Virtanen, "Heart rate monitors: State of the art," *Journal of Sports Sciences,* vol. 16, pp. 3-7, 1998/01/01 1998.

[7] T. Shany, S. J. Redmond, M. R. Narayanan, and N. H. Lovell, "Sensors-based wearable systems for monitoring of human movement and falls," *IEEE Sensors,* 2012 (preprint, accepted 10/4/2011).

[8] M. R. Narayanan, S. J. Redmond, M. E. Scalzi, S. R. Lord, B. G. Celler, and N. H. Lovell, "Longitudinal falls risk estimation using triaxial accelerometry," *IEEE Transactions on Biomedical Engineering,* vol. 57, pp. 534-541, 2010.

[9] S. S. S. Ng, T.-O. Chan, K.-W. To, J. Ngai, A. Tung, F. W. S. Ko, and D. S. C. Hui, "Validation of Embletta portable diagnostic system for identifying patients with suspected obstructive sleep apnoea syndrome (OSAS)," *Respirology,* vol. 15, pp. 336-342, 2010.

[10] S. J. Redmond and C. Heneghan, "Electrocardiogram-based automatic sleep staging in sleep disordered breathing," in *Computers in Cardiology*, Thessaloniki, Greece, 2003, pp. 609-612.

[11] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2 ed. New York: John Wiley & Sons, 2001.

[12] S. J. Redmond, N. H. Lovell, J. Basilakis, and B. G. Celler, "ECG quality measures in telecare monitoring," in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vancouver, Canada, 2008, pp. 2869-2872.

[13] S. J. Redmond, Y. Xie, D. Chang, J. Basilakis, and N. H. Lovell, "Electrocardiogram signal quality measures for unsupervised telehealth environments," *Physiological Measurement,* 2012 (in press, accepted 17/2/2012).

[14] Q. Y. Lee, G. S. H. Chan, S. J. Redmond, P. M. Middleton, E. Steel, P. Malouf, C. Critoph, G. Flynn, E. O'Lone, and N. H. Lovell, "Multivariate classification of systemic vascular resistance using photoplethysmography," *Physiological Measurement,* vol. 32, pp. 1117-1132, 2011.