Reconstruction of Missing Cardiovascular Signals using Adaptive Filtering

András Hartmann

Institute of Human Physiology and Clinical Experimental Research, Semmelweis University, Budapest, Hungary
INESC-ID, Lisbon, Portugal

Abstract

Here we introduce a robust method for filling in short missing segments in multiparameter ICU cardiovascular data inspired by the "PhysioNet/Computing in Cardiology Challenge 2010: Mind the Gap". Using the signals' history we identified the interconnections between the signals in the form of composite IIR transfer functions. Assuming that the connections do not vary in time, we managed to reconstruct the missing signals using the yet available parallel measured signals and the transfer functions. Since this assumption holds only for short timeperiod, we restricted the identification of the transfer function to segments prior the missing signal shorter than 30 seconds.

Our results are promising on the challenge dataset. We concluded that this approach can be efficient in reconstructing and even detecting missing or corrupted cardiovascular signals or other type of datasets with several modalities and strong interconnections between them.

1. Introduction

The physiological state of critically-ill Intensive Care Unit (ICU) patients can change frequently, demanding rapid analysis and quick decisions about interventions [1]. In this environment, where a continuous data flow of physiologic signals is indispensable, fault tolerance monitoring is of high importance. Missing or corrupted signals can occur among others due to human or machine error, sensor malfunction, moving artefacts and external noise. In ICU, several biomedical signals of interest are monitored parallel on the same patient with different modalities through different transducers. Although there is a fair amount of information that overlaps among these signals, resulting in high level of interconnection between them, the chance of artefacts affecting all the signals simultaneously is minor. There is thus a strong motivation to use advanced signal processing and machine learning methods that take advantage of the several different modalities to detect and/or eliminate artefacts [2], this has also inspired this year's PhysioNet/CinC challenge [3].

Here we introduce a robust batch processing method based on the short term static interconnections between cardiovascular signals measured parallel, that has proven useful in predicting 30 second segments of missing signals (gaps) in ICU datasets. The usage of the approach is not restricted to physiological data, but it is feasible in any application including multiple signals with static linear connections.

This paper is organized as follows: in Section 2 we introduce our method of identification and reconstruction, in Section 3 the results are described, and further discussed in Section 4. Finally in Section 5, conclusions are drawn and we point to future work directions.

2. Methods

In this work the Mind the Gap [3] dataset C was used, available at http://physionet.org [4]. The dataset consists of 100 ten-minute records containing 6, 7, or 8 signals acquired from bedside ICU patient monitors. The recorded signals vary across records, and they include ECG, continuous invasive blood pressure, respiration, fingertip plethysmograms, and occasional other signals. In one of these signals, the final 30-second segment (the target signal) is missing. Our goal was to reconstruct this missing target signal in each record, reaching possibly highest scores. Two types of scoring functions were introduced as

$$Q_1 = MAX\left(1 - \frac{MSE}{VAR}, 0\right) \quad (1)$$

$$Q_2 = MAX\left(CORR, 0\right), \quad (2)$$

being MSE the mean squared error, VAR the variance of the target signal, and $CORR$ the correlation coefficient between the target signal and the reconstruction. This scoring system is relevant because Q_1 measures the overall accuracy of the reconstruction, while Q_2 represents the accuracy in recovering the timing of the major fluctuations, which can be important at feature extraction, for example to derive RR interval tachogram from ECG recordings. As it can be seen on Eqs (1) and (2), both Q_1 and Q_2 scores range from zero to one.

2.1. The model

Our approach to identification was a gray-box technique. The model structure was derived from the principle that since the signals originate from the same physiological system, there has to be a strong interconnection between them. The connections were identified in the form of a Multi-Input/Single-Output (MISO) system. Third order Infinite Impulse Response (IIR) filters were used on the inputs, then the filtered signals were summarized in a linear model (See Figure 1).

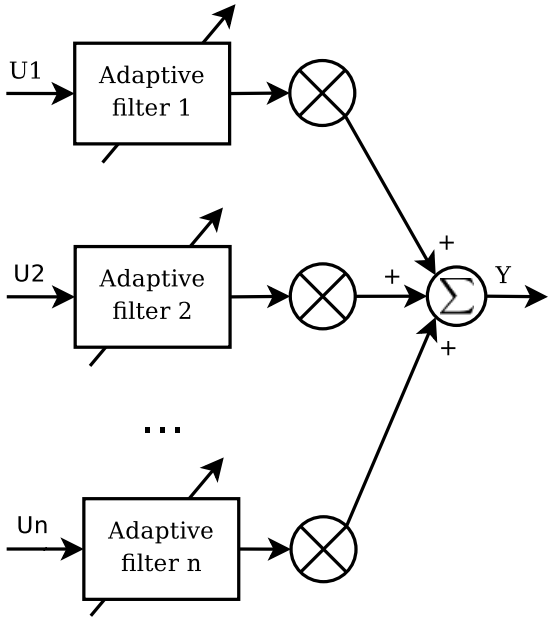

Figure 1. Model structure, where $U_1 ... U_n$ are the input signals and Y is the output.

2.2. The algorithm

The algorithm relies on the following assumptions:

1. There is a strong linear interconnection between the signals of the record.
2. Although there might be a time-variation in the connection between physiological signals, in short segments (<1 minute) this can be disregarded and the connection considered as time-invariant.
3. If we manage to identify a good model on the data available preceding the gap (prior), we will be able to reconstruct the target signal using the signals measured parallel to the target (concurrent signals).

The parameters of the model (the actual filter coefficients) were estimated using the prior as learning set. For a faster convergence, the mean was removed from the prior,

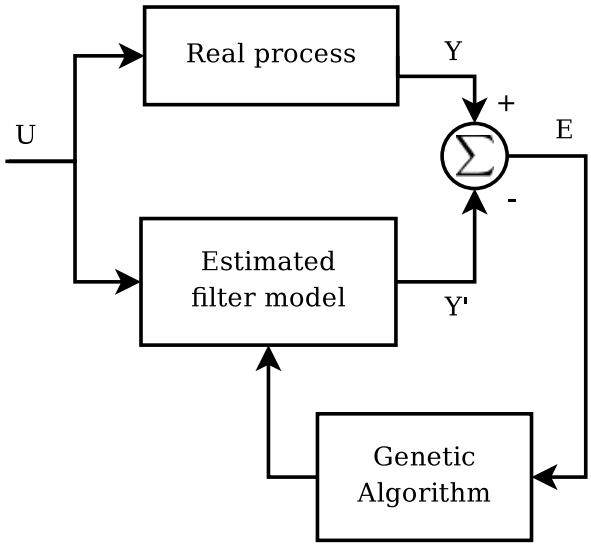

Figure 2. Block diagram of the identification method, where U is the input prior, Y is the history of the target signal, Y' is its estimate and E is an error for feedback to the learning algorithm.

and later added to the reconstructed signal. A genetic algorithm was used for parameter identification (see e.g. in [5]) with mean crossover strategy and mutation. The fitness function, $f = SD * MSE$ was found to maximize both Q_1 and Q_2 scores. The block diagram of the identification is shown in Figure 2.

Assuming that the connection is time-invariant, we were able to predict the target by using the identified model on the concurrent signals. In agreement with the 2nd assumption, the identification of the transfer function was restricted to the time interval 10 to 20 seconds preceding the target. The reconstructions fitted to different lengths were evaluated on the 30 second prior on a survival of the fittest basis. The algorithm was runned multiple times in order to overcome local optima.

A sample MATLAB implementation of the algorithm is freely available under the GNU Public License version 3 (GPLv3) from the author's homepage: http://researcherscorner.com/users/ahartmann.

3. Results

In all cases, the identified filters generated stable output. On most of the signals we found the presence of strong linear interconnection. A typical reconstruction and curve of parameter learning is presented in Figure 3.

Table 1 shows the detailed results grouped by the type of the target signals. As it can be seen, we resulted in an overall good reconstruction: $\sum Q_1 = 69.6591$ and $\sum Q_2 = 81.3236$ out of the possible 100. The Q_2 scores

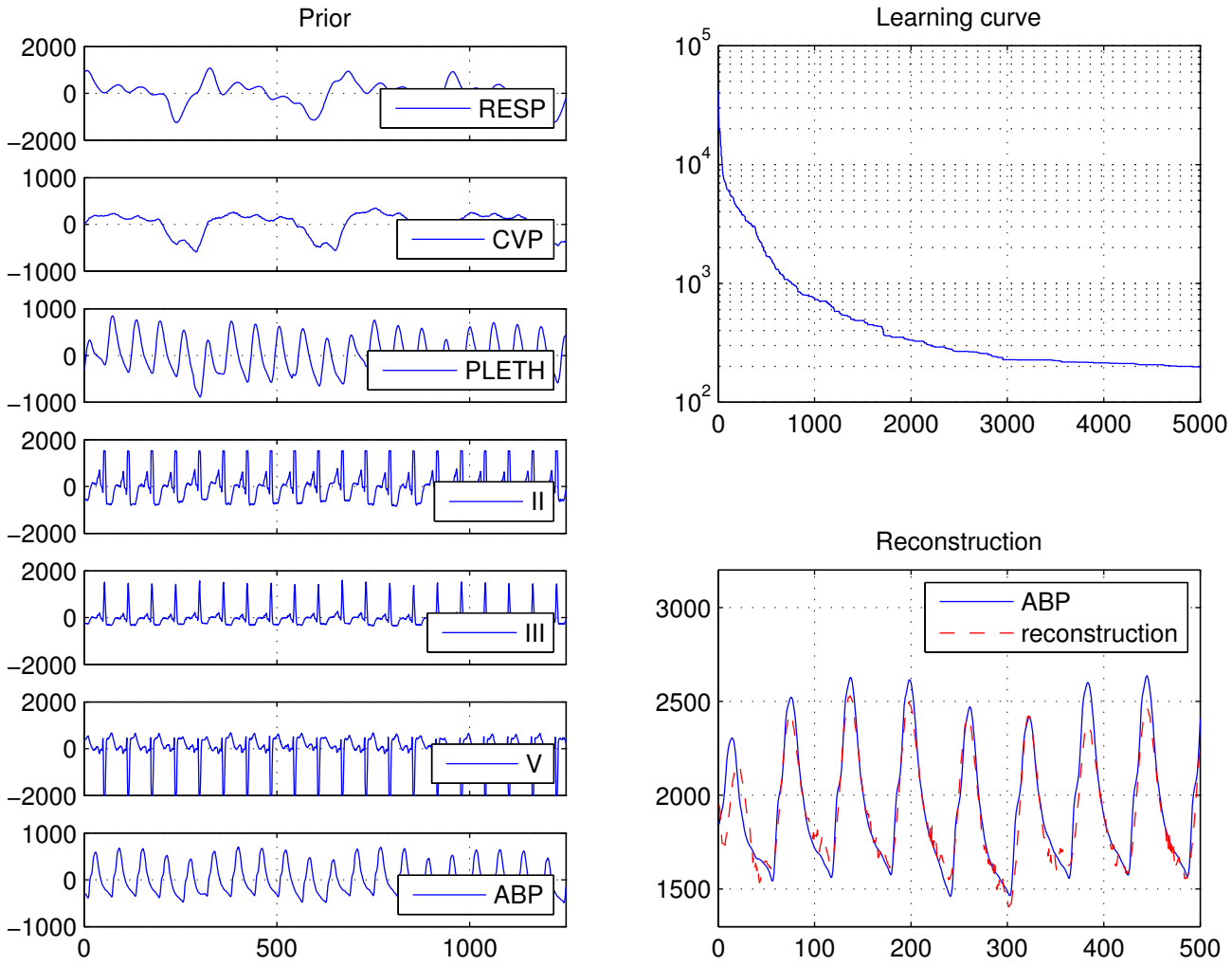

Figure 3. Demonstration of the (normalized mean) prior, parameter learning and reconstruction on a signal where the target was known. The datasource is a segment of the 'c21' record.

were found significantly higher than the Q_1 ones. The ECG signals resulted the best prediction, in some cases both Q_1 and Q_2 scores were over 0.995. The reconstruction of the further signals were usually considerably good, however on certain records, mostly when CVP, plethysmography or respiratory signals were missing, the reconstruction was moderate. We observed that the blood pressure signals had mostly better scores if another pressure signal was also available. It was also noticed, that the quality of the signals had an impact on the result.

The running time of the identification depends much on the length of the prior, the size of the population and the number of generations, which can be reduced by applying a condition to stop on satisfactory fitness. For example, the algorithm using 10 seconds of prior, a population of 100, and 2000 generations, the reconstruction of one signal takes about 12 minutes on an Intel(R) Xeon(R) E5310 1.60GHz CPU.

4. Discussion

The predictions show a good fit to the actual target, see Table 1. In agreement with [6] we presume that the identified transfer functions could potentially reflect the individual cardiovascular system. Note however, that it is not straightforward to order biological significance to the filter coefficients. Fortunately, for the purpose of reconstruction this is not necessary: all the knowledge about the system is abstracted in the parameters, allowing of the prediction independent from the domain of the original signals.

Table 1. Detailed scores and final result as mean±SD.

Target	Q_1	Q_2	Attempted
ECG	0.9 ± 0.13	0.95 ± 0.07	39
ABP	0.8 ± 0.13	0.9 ± 0.07	15
CVP	0.26 ± 0.31	0.44 ± 0.35	10
ICP	0.74 ± 0.35	0.93 ± 0.06	5
ART	0.49 ± 0.43	0.73 ± 0.23	3
PLETH	0.46 ± 0.34	0.62 ± 0.34	14
RESP	0.6 ± 0.27	0.76 ± 0.24	14
All	0.7 ± 0.31	0.81 ± 0.26	100

The results also confirm our preliminary assumptions (see section 2.2 for details). The fact that the ECG signals resulted the best reconstructions may be a consequence of every datafile containing at least two good quality ECG channels. We found the reconstructions of pressure signals to be closer to the target when another pressure signal was also available. These findings imply that the connection between the same modalities tend to be more significant than between different ones. Hence, a great opportunity to improve the reconstruction of respiratory signals could be to add estimated respiratory derived from ECG signals (see [7] and references).

On certain signals however we did not achieve a good reconstruction. The reason of this can be that there was no strong connection between the signals in the prior, or the connection was highly non-linear, which could not be captured well by our model. The score of the reconstruction is also highly influenced by the quality of the original signals. For instance if the target signal is badly scaled and exceeds the measurement range, the scores might be low even if the algorithm was able to predict the details where the actual target shows constant minimum or maximum values.

Theoretically, this approach is capable not only to fill in the gaps of known positions, but also to detect the points where the connections change drastically indicating corrupted signal. To be useful in clinical practice however, an on-line implementation would be necessary, even if detecting artefacts can potentially be of interest in batch processing of retrospective physiological data as well.

5. Conclusion and future work

The introduced identification algorithm is based on the strong linear connection between the parallel measured signals. In practice on short segments of cardiovascular records the connections proved to be time-invariant. By identifying these in batch processing, it was possible to reconstruct the missing target signal. The algorithm is not restricted to physiological signals, indeed it could potentially be used in many real-world applications.

Our future direction points towards reducing the running time of the identification and potentially also provide an algorithm feasible for on-line processing. We are also considering to include non-linear filters in the model, which may result a better identification and prediction of the underlaying signals.

Acknowledgements

The author would like to acknowledge the many valuable suggestions made by Lászlo Kocsis and Ana C. Mendes.

References

[1] Clifford GD, Long WJ, Moody GB, Szolovits P. Robust parameter extraction for decision support using multimodal intensive care data. Philosophical transactions Series A Mathematical physical and engineering sciences 2009; 367(1887):411–29. ISSN 1364-503X.

[2] Takla G, Petre J, Doyle D, Horibe M, B. The problem of artifacts in patient monitor data during surgery: a clinical and methodological review. Anesthesia amp 2006;103:1196–1204.

[3] Moody G. The PhysioNet/Computing in Cardiology Challenge 2010: Mind the Gap. In Computing in Cardiology (CINC), Belfast, volume 37. Belfast, 2010; [in press].

[4] Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov P, Mark R, Mietus J, Moody G, Peng C, Stanley H. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101(23):e215.

[5] Goldberg DE. Genetic Algorithms in Search, Optimization and Machine Learning. 1st edition. Addison-Wesley Longman Publishing Co., Inc., 1989. ISBN 978-0201157673.

[6] Chew C, Zahedi E. Limb Cardiovasculature System Identification Using Adaptive Filtering. In 3rd Kuala Lumpur International Conference on Biomedical Engineering 2006 IFMBE Proceedings, volume 15 of IFMBE Proceedings. Kuala Lumpur: Springer Berlin Heidelberg. ISBN 978-3-540-68016-1, 2007; 406–409.

[7] O'Brien C, Heneghan C. A comparison of algorithms for estimation of a respiratory signal from the surface electrocardiogram. Computers in biology and medicine 2007; 37(3):305–14. ISSN 0010-4825.

Address for correspondence:

András Hartmann
Institute of Human Physiology and Clinical Experimental Research, Semmelweis University
Pf. 448, H-1446
Budapest, Hungary
andras.hartmann@eok.sote.hu