A Probabilistic Beat-to-Beat Filtering Model for Continuous and Accurate Blood Pressure Estimation

1st Zehua Chen  
Department of EEE  
Imperial College London  
London, UK  
zehua.chen18@imperial.ac.uk

2nd Bruno Scalzo Dees  
Department of EEE  
Imperial College London  
London, UK  
bruno.scalzo-dees12@imperial.ac.uk

3rd Danilo P. Mandic  
Department of EEE  
Imperial College London  
London, UK  
d.mandic@imperial.ac.uk

Abstract—Cuffless technologies provide a convenient platform for remote and continuous blood pressure (BP) monitoring, however, signal recordings employed for cuffless BP estimation, which are based on the electrocardiogram (ECG) and photoplethysmography (PPG) signals, are frequently corrupted with measurement noise and artefacts. Consequently, even a small portion of abnormal data samples can severely impact the overall signal quality and therefore lead to significantly distorted BP value estimates. To this end, a data-driven model is proposed to infer the beat-to-beat signal quality for the ECG, PPG and BP signal recordings, whereby high-quality and low-quality (outlier) beats are detected using a probabilistic model chosen according to the maximum entropy principle. Physiological rules are also imposed to guarantee that each filtered sample is physiologically meaningful. The advantages of the proposed filtering framework for both systolic blood pressure and diastolic blood pressure estimation are demonstrated through the analysis and estimation of 12,000 clinical BP recordings, consisting of over 200,000 test samples.

Index Terms—Continuous Blood Pressure Estimation, Beat Quality Detection, Probabilistic Filtering Model, Systolic Blood Pressure, Diastolic Blood Pressure

I. INTRODUCTION

In 2015 alone, Cardiovascular Diseases (CVDs) have caused over 17.9 million deaths (32.1% of global death rate), thereby being the leading contributor to the global death rate [1]. In particular, high blood pressure has been responsible for 13% of these CVD-related deaths. Evidently, there is an urgent need for convenient and continuous blood pressure measurement technologies which can be utilised by patients in-home. The traditional approach to measuring blood pressure is based on the mercury sphygmomanometer which employs an inflatable cuff to collapse and release the artery beneath the cuff [2]. While cuffs prevent the BP measurement from being convenient and continuous, cuffless blood pressure estimation methods provide a solution platform for such needs.

In order to accurately estimate systolic Blood Pressure (SBP) and diastolic Blood Pressure (DBP) values, research from recent decades has focused on developing features for accurately estimating SBP and DBP values. For instance, a pulse transit time (PTT) model was introduced in [3] to estimate blood pressure from electrocardiogram (ECG) and photoplethysmography (PPG) signal recordings. The proposed features can be categorised into the following classes: (i) the inter-channel time features relevant to PTT [4], [5]; (ii) the waveform features from PPG signal [6], [7]; and (iii) the point features from the acceleration waveform of PPG signal (APG) [8]. While many features for blood pressure estimation have been proposed, their effectiveness is typically verified using a small dataset collected in ideal laboratory settings. The subjects involved to record the dataset are often healthy and the number of subjects is usually insufficient for statistically significant results. In turn, when these features are tested on larger datasets collected from patients in clinical environments, the detection uncertainty of these features usually increases as signals recorded in clinical environments are much noisier than those collected in laboratory settings. For instance, the noise, artefacts and missing data inherent to recordings performed in the intensive care unit (ICU) often corrupt the physiological recordings and therefore lead to significant errors in the physiological parameter estimates, such as the heart rate (HR). In order to reduce the ICU false alarm caused by HR estimate errors, Signal Quality Indexes (SQIs) have been developed for ECG, PPG and ABP signals. These SQIs are used to detect high- and low-quality beats in a signal recording, whereby low-quality beats are discarded so as to reduce the estimation error. These SQIs are typically constructed using flexible rules [9], time-domain and frequency domain features [10] and adaptive template matching [11].

Inspired by the SQI-approach to HR estimation error reduction, we propose a probabilistic data-driven framework for the detection of high- and low-quality (outlier) ECG, PPG and Arterial Blood Pressure (ABP) beats. As shown in Fig. 1, in the blood pressure (BP) estimation framework the main contribution of this work focuses on the pre-processing part. The BP estimation accuracy is expected to be improved through filtering out the low-quality ECG, PPG and ABP beats. Our proposed filtering model is unsupervised in the sense that it does not require a training (calibration) procedure for determining the optimal filtering thresholds. Also, without any requirement on manually defining and calculating features for assessing signal quality, this model utilizes raw signal to recognize high-quality beats. In addition, unlike the template matching method demanding an initial window to extract the average QRS complex or PPG/ABP pulse-waveform template, the proposed model is capable of evaluating beats in short
recordings, e.g., only a few seconds. Furthermore, the filtering model also employs physiological rules to ensure that filtered beats are physiologically meaningful. The advantages of the proposed filtering framework is demonstrated on real-world clinical recordings and is shown to effectively reduce the blood pressure estimation error.

II. SIGNALS AND DATABASE

The Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC) II dataset [12] is considered in this work for cuffless blood pressure estimation from ECG and PPG signals. We consider the version 3 of this dataset, published in UCI Machine Learning Repository [5]. The dataset consists of 12,000 ECG, PPG and ABP recordings, including: (i) ECG signals from channel II; (ii) PPG signals from the fingertip; and (iii) continuous invasively measured ABP recordings. The recordings were collected from ICUs at a sampling frequency of 125 Hz, whereby the time length of each recording ranged between 8s and 592s. Owing to the lack of patient information, e.g. the patient ID, we did not consider the impact of patient-specific recording discrepancies during the modelling and estimation procedures. It is also important to note that there exists unknown inter-channel delays (of up to 500 ms) and/or unspecified filtering delays, as mentioned in the technical limitations section of the MIMIC III WAVEFORM DATABASE [13]. Therefore, the well-known and widely-used feature PTT, which is measured between ECG and PPG signals, may be unreliable either in absolute or relative terms.

We employed a sliding window of 2s in length, with a step-size of 1s, for both the PPG feature extraction and blood pressure estimation. The detected maximum and minimum values of the blood pressure waveform in each window were first averaged and then taken as the SBP and DBP value of that window. Although this dataset, and the MIMIC II database, have been employed by a broad community of researchers, they typically consider only a small subset of the dataset. Although numerous reasons for discarding recordings have been described, no mathematical models for doing so have yet been considered. This would be highly desirable as discarded recordings may contain clean beats, and therefore useful information for BP value estimation. To this end, we develop a data-driven beat-to-beat signal quality assessment method, whereby features from the high-quality beats are used for the estimation procedure, while the features from the low-quality (outlier) beats are discarded.

Fig. 1: The Flowchart of This Study.

Fig. 2: Beat isolation.

III. MODEL

Fig. 1 demonstrates the flowchart of this study. Our main contribution focuses on the pre-processing part. In this section, our proposed probabilistic filtering model is discussed in detail.

A. Beat Isolation

Each cardiac cycle consists of two stages: (i) the heart muscle relaxes and refills with blood; (ii) the heart contracts and pumps blood. Throughout the cardiac cycle, the ECG, PPG and ABP signal value increase and decrease accordingly, thereby reflecting respectively the electrical activity of the heart, the blood volume changes and the pressure of blood circulating against the walls of the blood vessels. For each of these three signals, we isolate single beats, each of which ranges from the ending of a heartbeat to the beginning of the next. To perform the beat isolation, it is necessary to precisely locate the troughs of each signal. We therefore determine the minimum time interval between troughs for each signal, defined as \( T_{\text{min}} = \frac{1}{f_{\text{max}}} \), where \( f_{\text{max}} \) denotes the frequency (in Hz) associated with the maximum amplitude of the power spectral density (PSD) of the signal. In the absence of noise, \( T_{\text{min}} \) is equal to cardiac cycle.

To reduce the impact of noise specific to each signal type, we estimate the minimum interval, \( T_{\text{min}} \), using the weighted scheme

\[
T_{\text{min}} = \frac{1}{w_{\text{ECG}}f_{\text{ECG}} + w_{\text{PPG}}f_{\text{PPG}} + w_{\text{ABP}}f_{\text{ABP}}}
\]

where, \( w_{\text{ECG}} \in [0,1] \) denotes the confidence weighting assigned to the ECG signal. Without prior knowledge about the signal quality of ECG, PPG and ABP signals, we set \( w_{\text{ECG}} = w_{\text{PPG}} = w_{\text{ABP}} = \frac{1}{3} \) for simplicity. Fig. 2 shows the three signals extracted from recording 3 in the dataset, with the onset time being \( t_s = 90 \) s and the end time being \( t_e = 110 \) s. In this work, the peaks in each ECG, PPG or ABP recording are estimated by employing the MATLAB function \textit{findpeaks}. As illustrated, most beats are accurately isolated based on their detected peak (green marker) and trough (red marker). However, owing to the presence of noise and outliers, there are regions of the data with erroneously detected peaks and troughs. To rectify these errors, we next introduce our probabilistic beat-to-beat filtering model.

\[
T_{\text{min}} = \frac{1}{w_{\text{ECG}}f_{\text{ECG}} + w_{\text{PPG}}f_{\text{PPG}} + w_{\text{ABP}}f_{\text{ABP}}}
\]
B. Beat Normalization

The individual beats extracted from the raw input signals may be of different length and scale, thus it is necessary to normalize them so as to allow for their direct comparison, thereby mitigating the possible effects of their inherent non-stationarity during the filtering, feature creation and modelling procedures.

By denoting the $t$-th (time) value of the $n$-th beat by $x_n[t] \in \mathbb{R}$, the values are scaled as follows:

1. $x_n[t] \leftarrow \frac{x_n[t] - x_n[T]}{\max_{t} \{x_n[t]\} - x_n[T]}$
2. if $t \geq \arg \max_{t} \{x_n[t]\}$ then $x_n[t] \leftarrow \frac{x_n[T] - x_n[t]}{x_n[T] - \max_{t} \{x_n[t]\}}$

where $x_n[T]$ is the value of last point in each beat. In this way, the beats are scaled so that the first and last values equal 0, i.e. $x_n[1] = x_n[T] = 0$, while their maximum value equals 1, i.e. $\max_{t} \{x_n[t]\} = 1$. The result of the scaling procedure is displayed in Fig. 4 (c). Notice that the beats become homogeneous and are therefore suitable for modelling purposes.

To deal with beats of different sample length, we set their lengths to be equal to the median beat length, $T_m$, whereby beats shorter than $T_m$ are padded with zeros, and beats longer than $T_m$ are clipped.

C. Probabilistic Model

We next consider establishing an appropriate probabilistic model to describe the observed distribution of $x_n[t]$ at a given time instant, $t$, across all beats (see Fig. 3). In situations of limited knowledge, it is natural to choose the model according to the maximum entropy principle [14]. Such a model can be thought of as the most forgiving one as it allows for the largest magnitude and frequency of low probability events, i.e. outliers, thereby maximising entropy.

The model we consider assumes that the observation values, $x_n[t]$, consist of a mixture of frequent homogeneous (high-quality) signals and infrequent outlier (low-quality) signals. Realistic conditions we impose on the distribution include:

(i) the mean value of $x_n[t]$ across all beats is given by $\mu[t]$;
(ii) the values of $x_n[t]$ are absolutely summable over all $N$ beats at a given time instant, $t$, i.e. $\sigma[t] = \frac{1}{N} \sum_{n=1}^{N} |x_n[t] - \mu[t]| < \infty$.

The maximum entropy is thereby attained by the Laplace distribution [15], [16], i.e. $x_n[t] \sim \mathcal{L}(\mu[t], \sigma[t])$. Given the $t$-th value of $N$ i.i.d. beats, $x_{1}[t], x_{2}[t], \ldots, x_{N}[t]$, the maximum likelihood estimators of $\mu[t]$ and $\sigma[t]$ are respectively the
sample median and the mean absolute deviation from the sample median [17]:

High-quality beats are therefore located close to the median, $\mu[t]$. In turn, their distribution can be modelled according to a Gaussian located at the sample median, i.e. $x_n[t] \sim \mathcal{N}(\mu[t], \sigma[t])$. This model is suitable since the Gaussian exhibits thinner tails than the Laplace and therefore expects less outliers, shown in Fig. 3.

The likelihood of the sample $x_n[t]$ being a high-quality beat is therefore given by

$$p(x_n[t]) \propto \exp \left( -\frac{|x_n[t] - \mu[t]|^2}{2\sigma[t]^2} \right) \tag{2}$$

By considering the likelihood of $x_n[t]$, for $t = 1, \ldots, T_m$, the likelihood of the $n$-th beat being a high-quality one is given by

$$score_n = \exp \left( \frac{1}{T_m} \sum_{t=1}^{T_m} \log p(x_n[t]) \right) \tag{3}$$

The beats were then classified to be of high-quality according to the threshold, $score_n \geq \exp(-2)$, whereby, on average, the beat samples satisfy $\frac{|x_n[t] - \mu[t]|}{\sigma[t]} \leq 2$. From extracting raw signal to calculating the ratio of high- and low-quality beats, each step and corresponding results have been shown in Fig. 4.

Remark 1: The same scale parameter value, $\sigma[t]$, is employed for the Laplace and Gaussian models since, owing to Jensen’s inequality, the standard deviation is bounded from below by the mean absolute deviation, $E[|x_n[t] - \mu[t]|] \leq \sqrt{E[(x_n[t] - \mu[t])^2]}$.

D. Physiological Rules

In order to remove beats which are not physiologically meaningful/possible, we also impose a filtering stage based on physiological rules, thereby guaranteeing that each filtered beat is physiologically meaningful and possible. Such errors may arise due to incorrect measurements or recording artefacts. The range of the physiological rules is flexible and can be set to a probabilistically probable range for an adult population. If a beat does not satisfying all of the conditions, the beat is classified as an outlier beat and is therefore discarded.

In this work, we select three indices for ABP and two indices for both ECG and PPG. The pulse pressure (PP) is defined as the difference between SBP and DBP, while the RR interval is defined as the time interval between two adjacent ECG R-peaks or PPG pulse peaks. Similar abnormality criteria are employed for these indices in [9], [18]. The five physiological rules for are shown in Table I.

**TABLE I: Physiological rules.**

<table>
<thead>
<tr>
<th>Signal</th>
<th>Indices</th>
<th>Normal Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>SBP</td>
<td>SBP&lt;300/mmHg</td>
</tr>
<tr>
<td></td>
<td>DBP</td>
<td>DBP&gt;20/mmHg</td>
</tr>
<tr>
<td></td>
<td>PP</td>
<td>PP&gt;20/mmHg</td>
</tr>
<tr>
<td>ECG/PPG</td>
<td>HR</td>
<td>20bpm&lt;HR&lt;220bpm</td>
</tr>
<tr>
<td></td>
<td>RR</td>
<td>RR&lt;3s</td>
</tr>
</tbody>
</table>

E. PPG Features Detection

To demonstrate the performance of our filtering framework, we employ the classical and simple blood pressure estimation features which have been presented and verified effective by other researchers. Owing to the unreliable quality of the PTT features in this dataset, we do not use them in our analysis. For both the SBP and DBP estimation procedures, we employ 32 time-domain features extracted from each beat in the PPG signal, defined in Fig. 5. For clarity, these features are plotted with three continuous PPG beats in Fig. 5 while in real computation they are detected from every beat in PPG signal.

In detail, these 32 time-domain PPG features include: 1) **HR**: HR is estimated by the time distance between two continuous systolic peaks [5]. 2) **PPG Intensity Ratio (PIR)**: PIR can be used to assess the arterial diameter change during one cardiac cycle from systole to diastole thus being relevant to peripheral resistance and blood volume [19]. In calculation, PIR is detected as the ratio of PPG peak intensity $I_H$ to PPG valley intensity $I_L$ of one cardiac cycle. 3) **Dicrotic Notch (DN)**: Two DN features $V_{DN}$ and $T_{DN}$ [20] are employed here. $V_{DN}$ is the amplitude of DN while $T_{DN}$ is the time distance from onset to DN in each cardiac cycle. 4) **Large Artery Stiffness Index (LASI)**: LASI is linked with arterial stiffness [21]. It is calculated by the inverse of the time distance between systolic peak and DN. 5) **Systolic time and Diastolic time**: $T_S$ is the time distance from onset to systolic peak and $T_D$ is the time distance from systolic peak to end point. 6) **Maximum Slope**: $V_{MS}$ is the amplitude of the point with maximum slope within $T_S$ and $T_{MS}$ is the time distance from onset to that point. 7) **Area Ratio**: These four are the ratios of areas under PPG waveform between selected points, which is denoted by $S_1$, $S_2$, $S_3$ and $S_4$ here. 8) **Mean Value**: The mean value of whole PPG pulse in one cardiac cycle, the mean value of PPG pulse within $T_S$, the mean value of PPG pulse before DN and the mean value of PPG pulse after DN are also important BP estimation features [20]. 9) **Diastolic Arc Length**: The PPG curve length from systolic peak to end point in one cardiac cycle. 10) **Systolic Width (SW)** and **Diastolic Width (DW)**: The width of 10%, 25%, 33%, 50%, 66%, 75%, 90% in systolic time and diastolic time.

IV. RESULTS

We use a multi-layer neural network for both of the SBP and DBP estimation procedures. The structure contains one input layer, four hidden layers and one output layer, which presents the best prediction performance after experimenting a variety of neural network structures.

**TABLE II: Ratio of retained/discarded samples.**

<table>
<thead>
<tr>
<th>Filter Type</th>
<th>Retained(%)</th>
<th>Discarded(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilistic Model</td>
<td>86.41</td>
<td>13.59</td>
</tr>
<tr>
<td>Physiological Rules</td>
<td>98.87</td>
<td>1.13</td>
</tr>
<tr>
<td>In Total</td>
<td>85.52</td>
<td>14.48</td>
</tr>
</tbody>
</table>
We employ a moving-average window of 2 s in length, with a step-size of 1 s, to extract the 32 PPG features, the SBP value and the DBP value in each window. For clarity, and without loss of generality, we refer to each window as a sample. By employing the proposed filtering model, only the features extracted from the high-quality beats in each window are used to produce a sample. If none of the beats in a window are of high quality, the entire sample would be discarded. For the 12,000 recordings, 2,657,513 samples were extracted in total. Upon employing the proposed filtering model, 2,272,617 filtered samples were retained. For comparison purposes, the same features were also extracted without the filtering stage, whereby the features from each beat in a window were averaged and then taken as a sample. If none of the beats were detected in a window, this window would be discarded as well. Without our filtering model, 2,635,393 raw samples were retained.

Upon employing the proposed probabilistic filtering framework, the proportion of samples retained and discarded were recorded and are shown in Table II. In practice, the probabilistic modelling and physiological rules-based filtering were applied sequentially for each sample. After the filtering stage, we used the 90 % of the samples to form the training dataset and the remaining 10 % of the samples to form the test dataset. The 10-fold cross validation was employed here. In order to avoid information leakage, we separated the training dataset and test dataset according to the Recording ID. In this dataset, each patient may have more than one recording but the recordings of the same patient appear next to each other.

It is insightful to first illustrate the distribution of SBP and DBP in this dataset. The histogram of SBP and DBP samples are shown in Fig. 6, indicating the more centered nature of the distribution of the latter.

This paper realises a calibration-free, reliable and only PPG-based continuous BP estimation model. The mean error (ME), standard deviation (STD) and mean absolute error (MAE) of estimation results are used to evaluate the regression results. Table III demonstrates the increase of the BP estimation accuracy when the proposed probabilistic filtering framework is employed in comparison with using the raw data. Observe that with our filtering model, both the SBP and DBP estimation error is reduced, based on the error ME, MAE and STD measurement.

Also, Table III shows the comparison between this work and other research papers about continuous BP estimation. Both of the work from [22] and [19] only used relatively small sample size dataset and realised estimation with the need of calibration. In this work, the SBP and DBP estimation models, with and without the proposed probabilistic filtering, did not require any calibration. As the individually calibrated model will significantly improve the regression results, we mainly compare our work with [5] and [21] which also finished calibration-free BP estimation using similar public dataset. The dataset published by [5] includes 12,000 recordings extracted from MIMIC II database. In both [5] and [21], only a subset of these recordings was employed. However, some of the deleted recordings such as the ones with insufficient duration (defined as less than 10 minutes in [21]) should not be directly removed. In our work, all of the 12,000 clinical collected recordings were processed and used in training or test process. Using more real-world data, our results can be more statistically reliable. Moreover, because the MIMIC II website indicates the popular features PTTs are not reliable in that database, this work only used time-domain features from PPG signal. When PTT features are available in the future, our proposed model filtering out the low-quality beats in ECG and PPG would also be helpful to improve the reliability of them. The SBP and DBP estimation error would be potentially reduced with the use of PTT features.

V. CONCLUSIONS

Cuffless blood pressure measurement devices are necessary for convenient and continuous in-home monitoring of patients. However, data recorded in real life are likely to be corrupted by noise or measurement artefacts. We have therefore analysed the MIMIC II dataset so as to emulate the quality of the data collected in real-world clinical environments. A probabilistic filtering framework has been proposed which detects high- and low-quality (outlier) beats from ECG, PPG and ABP signal recordings. In this paper, in order to demonstrate the efficacy of the proposed filtering model, over 200,000 unseen samples have been employed and the error in both SBP and
DBP estimation tasks have been reduced in comparison with using the raw dataset. For improving estimation results in a further step, we plan to try transfer learning techniques and modifying the mean square error loss function. Furthermore, with key patient-specific features, such as the subject ID, age, sex, height, diseases or medicines, or if the PTT features can be employed, we might be able to further reduce the estimation uncertainty. Finally, our future work will include collecting our own recordings from patients and healthy subjects in order to further examine the performance of the proposed filtering model.

REFERENCES


