Classifying heart sounds using peak location for segmentation and feature construction

Elsa Ferreira Gomes
GECAD - Knowledge Eng.Decision Support
Institute of Engineering (ISEP/IPP)
Porto, Portugal

Emanuel Pereira
Institute of Engineering (ISEP/IPP)
Porto, Portugal

Abstract

In this paper we describe our methodology and results for the Classifying Heart Sounds PASCAL Challenge. We present our algoritm and results for segmentation and classification of S1 and S2 heart sounds.

1 Introduction

The goal of this challenge is to produce methods for Classifying Heart Sounds PASCAL Challenge. Specifically, the aim of this work is to create the first level of screening of cardiac pathologies both in a Hospital environment by a doctor (using a digital stethoscope) and at home by the patient (using a mobile device). The main components of heart sound signal of a normal heart are the first heart sound, S1 (or lub), corresponding to the systolic period, and the second heart sound, S2 (or dub), the diastolic period (Gupta *et.al*, 2007). In fact, this challenge is composed by two different challenges, Challenge1 and Challenge2. The aim of Challenge1, Heart Sound Segmentation, is to produce a method that can locate S1 and S2 sounds within audio data, segmenting the Normal audio files for two datasets, Datataset A and Dataset B. Dataset A contains data from the general public via the iStethoscope Pro iPhone app. Dataset B, contains data from a clinic trial in hospitals using the digital stethoscope DigiScope (Bentley *et.al*, 2011). These audio files are of varying lengths, between 1 second and 30 seconds (some have been clipped to reduce excessive noise and provide the salient fragment of the sound). The aim of Challenge2, Heart Sound Classification, is to produce a method that can classify real heart audio or beat classification into one of four categories for Dataset A (Normal, Murmur, Extra Heart Sound and Artifact) and three categories for Dataset B (Normal, Murmur and Extrasystole). Data is gathered in real-world situations and frequently contains background noise of every conceivable type. The differences between heart sounds corresponding to different heart symptoms can also be extremely subtle and challenging to separate. Success in classifying this form of data requires extremely robust classifiers. Despite its medical significance, to date this is a relatively unexplored application for machine learning (Bentley *et.al*, 2011). Here we present our work developed to answer to the proposed challenges, Challenge1 and Challenge2.

Preliminary work. Under review by AISTATS 2012. Do not distribute.

2 The challenges

As we said in the previous section, this challenge is, in fact, composed by two challenges. The goal of the first one is the heart sound segmentation and the goal of the second one is the heart sound classification. We know data have been gathered from two sources, resulting in two sets: Dataset A and Dataset B. We have different categories of data for each dataset. In the Normal category there are normal, healthy heart sounds. However, these sounds may contain a variety of background noises. They may also contain occasional random noise corresponding to breathing, or brushing the microphone against clothing or skin. A normal heart sound has a clear lub dub, lub dub pattern, with the time from lub to dub shorter than the time from dub to the next lub (when the heart rate is less than 140 beats per minute). We know most normal heart rates at rest are between about 60 and 100 beats (lub and dub) per minute. However, in these data, since the data may have been collected from children or adults in calm or excited states, the heart rates in the data may vary from 40 to 140 beats or higher per minute. In the Murmur category, the heart murmurs sound as though there is a whooshing, roaring, rumbling, or turbulent fluid noise in one of two temporal locations: (1) between lub and dub, or (2) between dub and lub. They can be a symptom of many heart disorders, some serious. One of the things that confuses non-medically trained people is that murmurs happen between lub and dub or between dub and lub; not on lub and not on dub. We also have the Extra Heart Sound category but only for Dataset A. These Extra heart sounds can be identified because there is an additional sound, e.g. a lub-lub dub or a lub dub-dub. An extra heart sound may not be a sign of disease. However, in some situations it is an important sign of disease, which if detected early could help a person. The extra heart sound is important to be able to detect as it cannot be detected by ultrasound very well. In the Artifact category, only existing on Dataset A, there are a wide range of different sounds, including feedback squeals and echoes, speech, music and noise. There are usually no discernible heart sounds, and thus little or no temporal periodicity at frequencies below 195 Hz. This category is the most different from the others. It is important to be able to distinguish this category from the other three categories, so that someone gathering the data can be instructed to try again. In the Extrasystole category, only existing in Dataset B, sounds may appear occasionally and can be identified because there is a heart sound that is out of rhythm involving extra or skipped heartbeats, e.g. a lub-lub dub or a lub dub-dub. Notice that an extrasystole may not be a sign of disease. It can happen normally in an adult and can be very common in children. However, in some situations extrasystoles can be caused by heart diseases. So, the files Atraining_normal.zip, Atraining_murmur.zip, Atraining_extrahs.zip, Atraining_artifact.zip, Aunlabelledtest.zip files were available for DatasetA. For Dataset B we had Btraining_normal.zip, Btraining_murmur.zip , Btraining_extrasystole.zip and Bunlabelledtest.zip.

3 Challenge1 - Heart Sound Segmentation

In the first challenge we aim to produce a method for determining the location of S1 and S2 sounds within audio data, segmenting the Normal audio files in Dataset A and Dataset B. The recorded signals were first preprocessed before performing segmentation. In the first step the signals were down sampled and filtered. In the second step, we have done the segmentation. Our algorithm is based on the envelope calculated using the normalized average Shannon energy (Liang, 1997).

3.1 Pre-processing

The original signal was decimated, using the decimate function of Matlab (MATLAB, 2009b) with factor 5. Then, a band-pass filter was applied. Considering the frequency components of S1 and S2 heart sounds, the chosen filter was a fifth order Chebyshev type I lowpass filter with cutoff from 100 Hz to 882 Hz. Then, the signals were normalized to the absolute maximum of the signal (Liang, 1997).

3.2 Segmentation

After pre-processing, we calculated the Shannon Envelope of the normalized signal. Then, the Average Shannon Energy is calculated in continuous 0.02 seconds windows throughout the signal with 0.01 second overlapping (Liang, 1997).

3.2.1 Peak finding

After obtaining the normalized average Shannon energy curve we identified the peaks. For that, we worked on the basis of the open source function peakdet, written in Matlab, with some changes made by us. This function finds the local maxima (and minima) using the strategy that a point is considered a maximum peak if it has a locally maximal value, and was preceded (to the left) by a value lower than a given delta. We have used two parameters of the function. The first is the vector to examine, and the second is the peak gap threshold (the delta). In our case considering this delta on the y values it was not enough. We also had to control the distance between peaks on the x-axis, because we know we cannot have two heart sounds too close. So, we have changed the function so that a local maximum is considered a peak if the distance to the nearest a peak is greater than a second threshold. Using this, we segmented almost all heart sounds. However, we also need to distinguish between S1 and S2. Our current approach for S1/S2 discrimination is still unsatisfactory. First, we tried to perform the detection of S1 and S2 sounds based on the fact that S2 is longer than S1, for normal heart rates (Kumar, 2006). Bearing this in mind we tried to pick each heart cycle and the corresponding systolic interval. The duration of S1-S2 sounds, or the distance between S1 and S2, was calculated and compared for every segment (Gupta *et.al*, 2007).The longest interval between two sounds was considered to correspond to the diastolic period and the sound at the right side was assigned as S1 and the sound at the left side was assigned as S2. Unfortunately, we find that those intervals vary widely from file to file, in our datasets. This happens because there are very different kinds of heart sounds data, for both datasets. We also tried to use a similar process to the energy, assuming S2 has higher frequency, but until now, we had no success. We are working on it.

3.3 Results

As we said before the task of the first challenge is to segment the audio files in Atraining_normal.zip and Btraining_normal.zip (for Dataset A and Dataset B, respectively). Our results were evaluated on a provided validation set with the correct locations of S1 and S2 sounds. This set contains the segmentation for sounds of the normal category from Dataset A and Dataset B. A test set for final evaluation was also available. This set contained hidden locations but provided the final evaluation results. In Figure 1 we can see the result of our method for the peak detection for file 103_1305031931979_D2.wav. In the 1st chart we have the original signal. In the 2nd chart we have the decimated signal. In the 3rd chart we have decimated signal filtered with a Chebyshev filter. In the 4rth chart we have the envelope of the signal and peaks. In the 5th chart we have the peaks over the original signal.

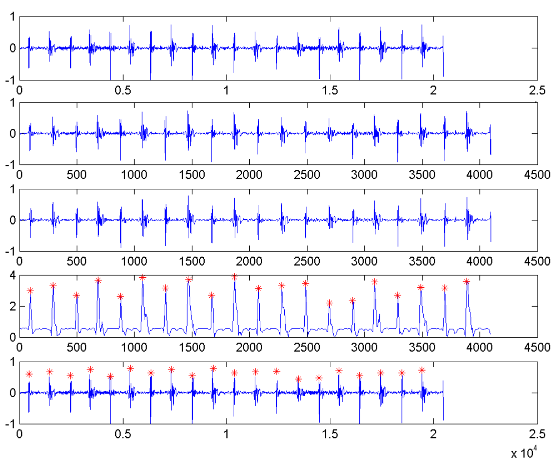

Figure 1: Peak detection on heart sound signal

In Table 1 we present our results for Dataset B. As we can see we have an error of 18.1 for file 175_1307987962616_B1 but an error 24149.9 for file 126_1306777102824_B. The total error was 72242.8. Some of these results could be better if we choose the first value for S1 or S2. However, because our method is not correctly implemented for the detection of S1 and s2 sounds, we have placed all first occurrences as S1. For example, for file 154_1306935608852_B1 if we start with a S2 the error decreases to 51.3 instead the error 2139.9 appearing in the table.

For Dataset A the error is higher then the error obtained for dataset B (see Table 2).

Table 1: Results for Dataset B

File Name	**Heart beats**	**Avg Error**
103_1305031931979_B	12.5	54.3
103_1305031931979_D2	10.0	35.7
106_1306776721273_B1	4.0	35.1
106_1306776721273_C2	3.0	27.3
106_1306776721273_D1	3.5	121.9
106_1306776721273_D2	8.0	4084.8
107_1305654946865_C1	7.5	1545.6
126_1306777102824_B	5.0	24149.9
126_1306777102824_C	3.0	13871.3
133_1306759619127_A	4.5	1578.2
134_1306428161797_C2	2.5	45.1
137_1306764999211_C	15.0	1629.0
140_1306519735121_B	11.0	49.4
146_1306778707532_B	18.0	2121.9
146_1306778707532_D3	3.0	26.8
147_1306523973811_A	3.5	3095.8
148_1306768801551_D2	5.0	10226.9
151_1306779785624_D	4.5	2560.3
154_1306935608852_B1	4.5	2139.9
159_1307018640315_B1	6.0	78.0
159_1307018640315_B2	3.0	66.7
167_1307111318050_A	13.0	58.0
167_1307111318050_C	3.5	1484.8
172_1307971284351_B1	3.5	68.4
175_1307987962616_B1	2.5	18.1
175_1307987962616_D	7.5	1813.2
179_1307990076841_B	16.5	63.2
181_1308052613891_D	3.0	40.0
184_1308073010307_D	26.5	70.7
190_1308076920011_D	4.0	1082.6

4 Challenge2 - Heart Sound Classification

The task of Challenge 2 is to produce a method that can classify the real heart audio into one of four categories for Dataset A (Normal, Murmur, Extra Heart Sound and Artifact) and three classes for Dataset B (Normal, Murmur and Extrasystole). This phase involves feature construction and selection and the goal of this phase of the challenge is correctly labeling the sounds in the provided files Aunlabelledtest.zip and Bunlabelledtest.zip

Table 2: Results for Dataset A

File Name	Heart beats	Avg Error
201101070538.aif	11.5	15324.8
201101151127.aif	1.0	516698.0
201102081152.aif	7.5	156064.2
201102201230.aif	9.5	88952.8
201102270940.aif	1.0	1445703.5
201103101140.aif	5.0	374151.6
201103140135.aif	5.0	314072.2
201103170121.aif	5.5	299314.2
201104122156.aif	2.0	640841.8
201106151236.aif	5.0	368613.5

4.1 Feature construction and selection

After the pre-processing and segmentation of the heart sound signal, some features were extracted. Currently, we are using six features (3), four of them were extracted from the distance between S1 and S2. Assuming that S1 corresponds to smaller segments and S2 to the others, we consider the ratio of the standard deviation of S1 over whole standard deviation and S2 ($Rs1$ and $Rs2$, respectively) and the ratio of the mean of S1 over the total mean and the ratio of S2 ($Rm1$ and $Rm2$, respectively). Another feature, $Rmedian$, is the ratio of the median of the three largest segments in the sample over total mean. The sixth feature, $R2$, is r square of the array of the sorted segments of the sample. After obtaining the features we used two different methods from the Weka data mining suite (Witten and Frank, 2005): J48, which generates decision trees, and MLP, the Multi Layer Perceptron.

Table 3: Features

Feature	Description
$Rs1$	$\sigma_{S1}/\sigma_{total}$
$Rs2$	$\sigma_{S2}/\sigma_{total}$
$Rm1$	μ_{S1}/μ_{total}
$Rm2$	μ_{S2}/μ_{total}
$Rmedian$	$median(max_1, max_2, max_3)/\mu_{total}$
$R2$	r square of the sorted segments

4.2 Results

In Challenge 2, the effectiveness of our classification approach is assessed using three metrics (per dataset) which are calculated from the tp (true positives), fp (false positives), tn (true negaties) and fn (false negatives) values. The metrics are precision per class, the Youden's Index, the F-score (only for Dataset A) and the Discriminant Power (only for Dataset B). Precision gives us the positive predictive value (the proportion of samples that belong in category c that are correctly placed in category c). Youden's Index has been used to compare diagnostic abilities of two tests, by evaluating the algorithm's ability to avoid failure. In Dataset A, Youden's Index is evaluated for Artifact category. In Dataset B the Youden's Index is calculated for the problematic heartbeats (Murmur and Extrasystole categories combined). Discriminant Power evaluates how well an algorithm distinguishes between positive and negative examples. It is a poor discriminant for a value ¡ 1, limited for a value ¡ 2, fair for a value ¡ 3, and good in other cases. Here we calculate the Discriminant power of problematic heartbeats (Murmur and Extrasystole categories combined).

In Table 4 and Table 5, we present the results for Dataset A and Dataset B, obtained after applying the J48 and MLP methods. As in Challenge 1, a testset was provided where we could test our method's effectiveness on the unlabelled set.

Table 4: Challenge 2 evaluation for DatasetA

	J48	MLP
Precision of Normal	0.25	0.35
Precision of Murmur	0.47	0.67
Precision of Extrasound	0.27	0.18
Precision of Artifact	0.71	0.92
Artifact Sensitivity	0.63	0.69
Artifact Specificity	0.39	0.44
Heartproblem Detection Sensitivity	0.55	0.45
Heartproblem Detection Precision	0.40	0.43
Youden Index of Artifact	0.01	0.13
F-Score of Heartproblem Detection	0.20	0.20
Total Precision	1.71	2.12

Table 5: Challenge 2 evaluation for DatasetB

	J48	MLP
Precision of Normal	0.72	0.70
Precision of Murmur	0.32	0.30
Precision of Extrasystole	0.33	0.67
Sensitivity of heart problem	0.22	0.19
Specificity of heart problem	0.82	0.84
Youden Ind Heartproblem Detection	0.04	0.02
Discriminant Power	0.05	0.04
Total Precision	1.37	1.67

As we can see in Table 5, our method has problems in classifying the non-normal heart beats, for Dataset B. In Dataset A the situation is different. In this case the normal class is one of the most difficult (Table 4). However, we think we can improve our method by improving the correct identification of S1 and S2 in the segmentation and by finding new features that take advantage of this identification.

5 Conclusions

In this paper, we present a methodology for the Classifying Heart Sounds PASCAL Challenge. We propose an algorithm for S1 and S2 heart sound identification (without ECG reference). The segmentation is accomplished by using the envelope of Shannon energy and an algorithm for peak detection. Despite of the good performance for the correct detection of S1 and S2 sounds in the signal, we need to improve the criteria for identifying S1 and S2 (who is who). After the segmentation, we used J48 and MLP algorithms (using Weka) to train and classify the computed features into Normal, Murmur or Extrasystole for Dataset B and Normal, Murmur, Extrasound and Artifact for Dataset A. We think this is a good basis for further analysis of the heart sound signals.

References

H. Liang, S. Lukkarinen, and I. Hartimo. (1997) Heart Sound Segmentation Algorithm Based On Heart Sound Envelogram. *Computers in Cardiology* **24**:105-108.

P. Bentley, G. Nordehn, M. Coimbra, and S. Mannor. The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results, www.peterjbentley.com/heartchallenge/ index.html.

D. Kumar, P. Carvalho, M. Antunes, P. Gil, J. Henriques, and L. Eug. (2006) A new algorithm for detection of s1 and s2 heart sound. IEEE International Conference on ICASSP 2006, pp. 1180-1183.

C. N. Gupta, R. Palaniappan, S. Rajan, S. Swaminathan, and S. M. Krishnan (2005) Segmentation and classification of heart sounds 1. *Canadian Conference on Electrical and Computer Engineering.* 1674-1677.

C. Gupta, R. Palaniappan, S. Swaminathan, and S. Krishnan. (2007) Neural network classification of homomorphic segmented heart sounds. *Applied Soft Computing* **7 (1)**:286-297.

E. Billauer,3.4.05 (2011), www.billauer.co.il/ peakdet.html

MATLAB version 7.9.0. Natick, Massachusetts.The MathWorks Inc., 2009b.

I. H. Witten and E. Frank (2005) Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.