

1 Abstract

One the widely used heart-condition examining method is phonocardiogram (PCG), a plot of high-fidelity recording of the sounds and murmurs made by the heart with the help of the machine called the phonocardiograph. By analyzing PCG record doctors can detect abnormal heart activity that classifies for special treatment. Such classification requires expert knowledge and can be very time consuming. By introducing programs for initial PCG classification we can reduce time spent on PCG analyzing.

From algorithmic point of view the problem described in this paper belongs to the well known classification problem category. Such issues are usually solved in three steps: segmentation, feature extraction and classification. Although classification step has been widely described, there are many good classifiers and approaches towards distinguishing elements in whole set, segmentation and feature extraction need more flexible approach. Every problem needs its own, individual solution. For the sake of the Physionet competition, segmentation has already been prepared, which uses state-of-the-art algorithms. Simple feature extraction is also provided withing the sample code, but it could be changed in order to gain better results.

The first step towards preparing solution will require finding the best classifier. We will compare effectiveness of some of the best known methods by taking their accuracy and speed into account. This will include SVM, Random Forests, KNN and HMM methods. Once their training has finished, computations for provided test sets will be held, and the results will be compared. It's worth noting that parameters for each classifier will have their values set according to grid search method results.

The second stage will consist of changing extracted features, adding new ones to enhance classification rates. As for now there are no specific ideas about what kind of features should be extracted or taken into account, although it's certain that they will be based on already implemented segmentation.

Samples used in this work are provided by Physionet challenge creators and are accessible during its duration. Whole records set consisted of 3126 files containing PCG recordings grouped into 6 folders. The validation package will be used as training set during classifiers' learning process. Audio files available within additional .zip file named *training* divide into 5 folders named: *training - a*, *training - b*, ..., *training - e* and will be used as test sets for which classifier results will be compared. All of mentioned here audio examples have already been classified by an expert and the results for their classification is known.