DropConnected Neural Network Trained with Diverse Features for Classifying Heart Sounds

Edmund Kay[1], Anurag Agarwal[1]

[1] Engineering Department, University of Cambridge, Cambridge, U.K.

Abstract

A fully-connected, two-hidden-layer neural network trained by error backpropagation, and regularized with DropConnect is used to classify heart sounds as normal or abnormal. The heart sounds are segmented using an open-source algorithm based on a hidden semi-Markov model. Features are extracted from the heart sounds using a wavelet transform, mel-frequency cepstral coefficients, inter-beat properties, and signal complexity. Features are normalized by subtracting by their means and dividing by their standard deviations across the whole training set. Any feature which is not significantly different between normal and abnormal recordings in the training data is removed, as are highly-correlated features. The dimensionality of the features vector is reduced by projecting it onto its first 70 principal components.

A 10 fold cross-validation study gives a mean classification score of 84.1% with a variance of 2.9%. The final score on the test data was 85.2%.

1. Introduction

An open-access database of heart sounds has been provided by Liu et al. [1]. Using this, an algorithm has been produced that classifies these heart sounds as normal or abnormal.

2. Data stratification

The make-up of the initial training data is given by Liu et al. [1]. The data removed from this to be used for validation is given in table 1. The remaining data is to be used for training. This data is balanced so that there are roughly equal numbers of normal and abnormal recordings in the training set. This is done by removing or repeating some recordings (with never more than 1 repeat of each recording). This final training data is given in table 2.

Set	N,G	N,P	A,G	A,P
a	40	0	37	3
b	44	5	36	13
c	3	0	3	0
d	5	0	4	1
e	49	0	38	15
f	8	0	8	0
Total	149	5	126	32

Table 1. Composition of validation data - N = normal, A = abnormal, G = good quality recording, P = poor quality recording

Set	N,G	N,P	A,G	A,P	Notes
a	152	2	239	13	Normals repeated
b	60	25	37	18	1/4 normals kept
c	8	0	16	3	Normals repeated
d	21	1	22	1	Nothing done
e	175	9	108	22	1/10 normals kept
f	70	2	46	6	Abnormals repeated
Total	486	39	468	63	

Table 2. Composition of final training data - N = normal, A = abnormal, G = good quality recording, P = poor quality recording

3. Segmentation

The heart sounds were segmented using Springer et al.'s open-source algorithm based on a hidden semi-Markov model [2].

4. Feature Extraction

Once the heart sounds have been segmented, features are extracted to best represent the heart sounds to the classifier.

4.1. Wavelet Transform

The wavelet transform is used to get a typical time-frequency representation of the sounds from a single heart cycle. The Morlet wavelet is used as the mother wavelet. The continuous wavelet transform (CWT) is evaluated at

11 frequencies, then each frequency level is individually normalized by subtracting by its mean and dividing by its standard deviation (across all time). Following this the CWT is averaged into 20 time bins per heart beat, 3 in S1, 7 in systole, 3 in S2, and 7 in diastole. This CWT is then averaged over heart beats which are well correlated with each other. This is done because, while recording the sound signal, stethoscope movements can lead to varying amplitudes, resulting in some parts of the cardiac cycle being artificially louder than others. The beats which are well correlated are determined by finding the two beats with the minimum euclidean distance between their CWT coefficients at each frequency. Then any other beats within 50% of this minimum euclidean distance, from both of the two beats, are also averaged over. This leads to a time-frequency representation of a typical heart beat which has 20 discrete points in time and 11 discrete points in frequency. A typical wavelet feature vector for a normal and abnormal recording are shown in figures 1 and 2 respectively. These figures show that, for the aortic stenosis patient (figure 2), between 150 and 250 Hz, the sound in the systolic phase has a higher amplitude relative to the fundamental heart sounds.
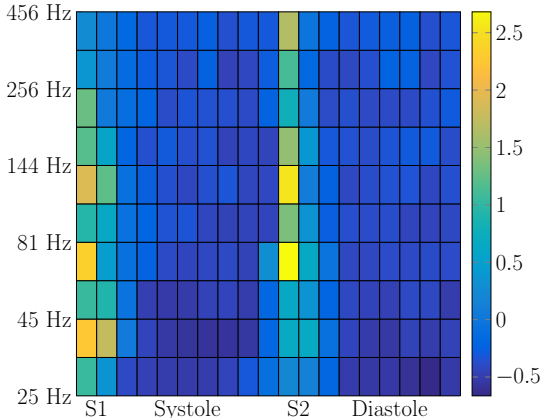

Figure 1. CWT feature for a normal heart sound (c0011)

4.2. Mel-frequency Cepstral Coefficients

Mel-frequency cepstral coefficients (MFCCs) have been widely used in speech recognition [3], and have also shown to be useful in heart sound classification [4]. In order to calculate the MFCCs, the signal is divided into the same 20 frames in time per heartbeat as for the CWT. Following this, a periodogram is found of each segment, with a 10% overlap either side. A Hamming window is used to reduce spectral leakage. The periodogram gives values for the signal power in 40 evenly spaced frequency bins between 0 (not inclusive) and the Nyquist frequency (inclusive). As for the CWT, this periodogram is then aver-
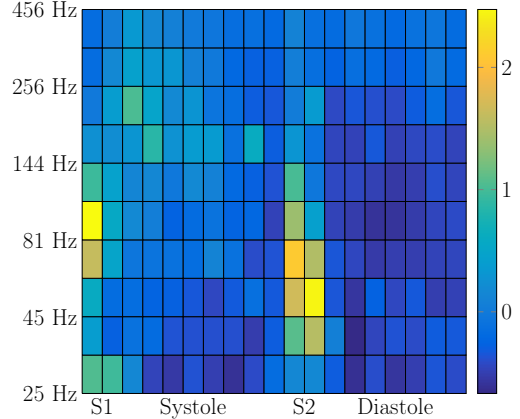

Figure 2. CWT feature for an abnormal heart sound (c0028 - aortic stenosis)

aged over beats which are well correlated with each other, at each frequency (see section 4.1 for full details). The periodogram is then filtered using a filter bank, in which frequencies are equally spaced in the mel-scale (equation 1).

$$Mel = 1125 \log\left(1 + \frac{f}{700}\right). \qquad (1)$$

The MFCCs are then obtained by taking the logarithm of each of the filtered periodograms, and then taking a discrete cosine transform (equation 2) of each of the 20 frames.

$$MFCC(t,k) = \sum_{n=1}^{N} \log\left(P_{filt}(t,n)\right) \cos\left(\frac{k\pi}{N}\left(n - 0.5\right)\right) \qquad (2)$$

$MFCC(t,k)$ gives the kth cepstral feature of the tth time frame. $P_{filt}(t,n)$ is the filtered power at time frame t for the nth filter bank. N is the number of filter banks used.

Finally, the last cepstral feature is removed to give the final MFCC feature vector used for classification.

4.3. Inter-beat features

The features, which were supplied for the challenge, were used. These are the mean and standard deviation of: the length of one heart cycle; the length of S1; the length of systole; the length of S2; the length of diastole; the ratio of systolic length to whole heart cycle length; the ratio of diastolic length to whole heart cycle length; the ratio of systolic length to diastolic length; the ratio of mean systolic amplitude to mean S1 amplitude; and the ratio of mean diastolic amplitude to mean S2 amplitude.

These features help to characterize the difference in behaviour between different heart beats in the cycle, which is missed by the first two features.

4.4. Complexity and Simplicity

Features which characterise the complexity of the signal are also extracted. These have also been used by Schmidt et al. [5]. Firstly a periodogram is found with 20 time frames per heart cycle (3 in S1, 7 in systole, 3 in S2, 7 in diastole) and 5 equally spaced frequency frames. This is obtained in the same way as for the MFCCs (section 4.2) and is then normalised between 0 and 1 by dividing by the largest value. Following this the spectral entropy (SE) is obtained as

$$SE\left(t\right) = \sum_f P_{xx}\left(t, f\right) log\left[P_{xx}\left(t, f\right)\right]. \quad (3)$$

Also the unbiased, standard deviation (SD), skewness (SK) and kurtosis (KT) of the power spectrum at each frequency, are obtained.

The spectral entropy, standard deviation, skewness, and kurtosis make up 35 features which describe the complexity of the signal.

5. Feature Normalization

Feature extraction results in a features vector of length 675. The make-up of the features vector is summarized by table 3.

Feature	CWT	MFCC	Inter-beat	Complexity
Length	220	400	20	35

Table 3. Make-up of the features vector

Features are then normalized by subtracting their means and dividing by their standard deviations (across the whole training set).

6. Feature Selection

After normalization, all features are subjected to a Student's t-test to determine whether they are significantly different between normal and abnormal recordings. Any feature with a test statistic less than the student's test statistic, from a two-tailed test at the 5% significance level, is removed.

Following this, for any pair of features which are highly correlated (a covariance greater than 0.9), one of them is removed (the one with the lowest t-statistic).

Finally, principal component analysis (PCA) is used to reduce the dimensionality of the features vector. The best performance (from the cross-validation study) was found when the features vector was projected onto its first 70 principal components.

7. Classification

The classification algorithm is based on a fully-connected, two-hidden-layer neural network, trained by error backpropagation [6]. The hyperbolic tangent activation function is used for all the neurons in the network except in the final layer, where the softmax activation function is used. The log-likelihood cost function is used. The range of hyper-parameters chosen for training the networks are given in table 4. In order to militate against overfitting,

Parameter	Value
Number of epochs	75 & 150
Mini-batch size	8
Learning rate	0.05,0.1 & 0.2
L2-regularization parameter	0 & 0.5
Momentum coefficient	0.3

Table 4. Parameters used for training neural network

two types of regularization are used. The first is L2 regularization, where a w^2 term is added to the cost function (where w is the weight along an individual neuron) to penalize large weights in the network. The second is DropConnect, which is described by Wan et al. [7]. For all the networks trained, the percentage of neurons which are randomly removed from each layer varies between 10 and 30% for the neurons between the input and the hidden layers, and between 30 and 70% for the neurons between the two hidden layers and between the second hidden and the output layers

8. Results

In order to estimate the performance accurately, a 10-fold cross-validation is performed. For this, the validation and training data are combined. Then for each cross-validation a tenth of this data is randomly selected for validation (preserving the ratio of recordings from each set), with the remaining data used for training. Any recordings that are repeated between training and validation are then removed from the validation set, for each cross-validation. The results of this 10-fold cross-validation are shown in table 5. The parameters which produced the networks with the highest 10-fold cross-validation accuracy are given in table 6. Table 7 shows the mean score on each of the six datasets across a 10-fold cross-validation.

9. Submission for test

When this algorithm was submitted to be evaluated on the test data, a number of different networks were trained with a range of hyper-parameters and different training sets. For example, a 5-fold cross validation is done with

Fold	Se /%	Sp /%	Sc (V) /%	Sc (T) /%
1	89.1	86.1	87.6	95.9
2	81.0	84.9	82.9	95.9
3	80.4	84.9	82.7	95.7
4	80.9	82.8	81.8	95.4
5	89.5	87.5	88.5	96.0
6	91.6	82.2	86.9	95.9
7	84.6	85.8	85.2	95.6
8	79.8	87.0	83.4	96.0
9	89.5	74.3	81.9	96.1
10	81.4	77.9	79.6	96.7
Mean	**84.8**	**83.3**	**84.1**	**95.9**
SD	4.6	4.2	2.9	0.3

Table 5. Results of 10-fold cross-validation test, where each score is the average score of 8 networks trained with the same hyper-parameters. Se = Sensitivity, Sp = Specificity, Sc = score, V = validation data, T = training data.

Parameter	Value
Number of nodes in hidden layers	35,28
Number of nodes in output layer	2
Number of epochs	150
Mini-batch size	8
Learning rate	0.05
L2-regularization parameter	0
Momentum coefficient	0.3
Percentage of neurons dropped I-1	20
Percentage of neurons dropped between 1-2	50
Percentage of neurons dropped between 2-O	50

Table 6. Parameters which achieve the best 10-fold cross-validation results, I-1 = between the input and first hidden layers, 1-2 = between the first and second hidden layers, 2-O = between the second hidden and output layers

Dataset	a	b	c	d	e	f
Score /%	71.5	77.9	80.0	77.5	94.0	73.5

Table 7. Mean score on each dataset from a 10-fold cross validation

networks trained on different hyper-parameters. The networks are then ensembled based on their score on the validation data and their diversity (measured by which recordings they incorrectly classified). Each of the networks in the ensemble classifies the heart sounds, with the final classification is given by the majority.

The best result obtained by this ensemble of networks on the test data was 85.2%.

10. Conclusion and Future Work

An algorithm capable of classifying heart sounds as normal or abnormal with an accuracy of 84% has been produced. A total of 675 features are extracted from a recording using a wavelet transform, mel-frequency cepstral coefficients, inter-beat properties, and signal complexity measures. These features are then normalized across the entire training data. Features are removed if they were shown to be statistically insignificant between normal and abnormal recordings. Highly correlated features, with a covariance greater than 0.9, are also removed. The dimensionality of the features vector is reduced by projecting it onto its first 70 principal components. Classification is done using a fully-connected, two-hidden-layer neural network, trained with error backpropagation and regularized using DropConnect. Networks trained with different hyper-parameters and trained on different data were then ensembled to produce the highest classification accuracy on the test data

Following this, an algorithm will be made which is capable of recognizing a good quality signal from a poor quality one. We will also try and produce an algorithm capable of diagnosing specific pathologies from the heart sounds.

References

[1] Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, Castells F, Roig JM, Silva I, Johnson AE, Syed Z, Schmidt SE, Papadaniil CD, Hadjileontiadis L, Naseri H, Moukadem A, Dieterlen A, Brandt C, Tang H, Samieinasab M, Samieinasab MR, Sameni R, Mark RG, Clifford GD. An open access database for the evaluation of heart sound algorithms. Physiological Measurement 2016;37(11).

[2] Springer D, Tarassenko L, Clifford G. Logistic Regression-HSMM-based Heart Sound Segmentation. IEEE Transactions on Biomedical Engineering March 2016;63(4):1–1.

[3] Huang X, Acero A, Hon HW. Spoken Language Processing. A Guide to Theory, Algorithm, and System Development. Prentice Hall, 2001.

[4] Wang P, Lim CS, Chauhan S, Foo JYA, Anantharaman V. Phonocardiographic Signal Analysis Method Using a Modified Hidden Markov Model. Annals of Biomedical Engineering December 2006;35(3):367–374.

[5] Schmidt SE, Holst-Hansen C, Hansen J, Toft E, Struijk JJ. Acoustic Features for the Identification of Coronary Artery Disease. IEEE Transactions on Biomedical Engineering October 2015;62(11):2611–2619.

[6] Bishop CM. Neural Networks for Pattern Recognition. Oxford University Press, November 1995.

[7] Wan L, Zeiler M, Zhang S, Cun YL, Fergus R. Regularization of Neural Networks using DropConnect. In International Conference on Machine Learning. 2013; 1–9.

Address for correspondence:

Name: Edmund Kay
Full postal address: Engineering Dept, Trumpington St, Cambridge, CB2 1PZ
E-mail address: ek360@cam.ac.uk