Yang et al. EURASIP Journal on Audio, Speech, and Music Processing (2016) 2016:9
DOI 10.1186/s13636-016-0086-9

EURASIP Journal on Audio, Speech, and Music Processing
a SpringerOpen Journal

RESEARCH Open Access

Semi-supervised feature selection for audio classification based on constraint compensated Laplacian score

Xu-Kui Yang[1], Liang He[2], Dan Qu[1*], Wei-Qiang Zhang[2] and Michael T. Johnson[3]

Abstract

Audio classification, classifying audio segments into broad categories such as speech, non-speech, and silence, is an important front-end problem in speech signal processing. Dozens of features have been proposed for audio classification. Unfortunately, these features are not directly complementary and combining them does not improve classification performance. Feature selection provides an effective mechanism for choosing the most relevant and least redundant features for classification. In this paper, we present a semi-supervised feature selection algorithm named Constraint Compensated Laplacian score (CCLS), which takes advantage of the local geometrical structure of unlabeled data as well as constraint information from labeled data. We apply this method to the audio classification task and compare it with other known feature selection methods. Experimental results demonstrate that CCLS gives substantial improvement.

Keywords: Audio classification, Semi-supervised feature selection, Locality preserving, Constraint information

1 Introduction

Initial classification of audio segments into broad categories such as speech, non-speech, and silence provides useful information for audio content understanding and analysis [1], and it has been used in a variety of commercial, forensic, and military applications [2]. Most audio classification systems involve two processing stages: feature extraction and classification. There is a considerable amount of literature on audio classification regarding different features [3] or classification methods [4]. Many features [5] have been developed to improve classification accuracy. Nevertheless, using all of these features in a classification system may not enhance but instead degrade the performance. The underlying reason is that there can be irrelevant, redundant, and even contradictory information among these features. Choosing the most relevant features to improve the classification accuracy is a challenging problem [6].

Feature selection methods can be divided into three categories: supervised, unsupervised, and semi-supervised. Supervised approaches require a large quantity of labeled data, and they are apt to ignore the internal structure of data by focusing too much on label information. Unsupervised feature selection fails to extract more discriminative features which may yield worse performance. Semi-supervised feature selection focuses on maximizing data effectiveness by using labeled and unlabeled data together [7]. In this case, the amount of unlabeled data is much larger than that of labeled data. Semi-supervised algorithms have attracted attention for their ability to model the intrinsic structure of data.

Approaches to feature selection are generally categorized into filter, wrapper, and embedded techniques. Filter methods use scores or confidences to evaluate the importance of features in the learning tasks and include algorithms such as Laplacian score (LS) [8], constraint score (CS) [9], and constrained Laplacian score (CLS) [10, 11]. Wrapper approaches evaluate different subsets of features and select the one with the best performance. The embedded model techniques search for the most relevant and effective features for models. The most common embedded methods are regularization-based [12], including LASSO, elastic net, or ridge regression. Since filter approaches can be applied to a broad

* Correspondence: qudanqudan@sina.com
[1]Zhengzhou Information Science and Technology Institute, Zhengzhou, China
Full list of author information is available at the end of the article

range of classification and learning methods, they have been widely used for their better generalization properties.

For audio classification, it is computationally challenging to evaluate the features' properties by testing them individually [13] or analyzing their characteristics individually [14]. Although some recent work on feature selection algorithms has focused on improving these weaknesses [15, 16], an efficient and effective method has yet to be developed. This is primarily because most approaches rely on labeled data. It is hard to get sufficient labeled data for the evaluation of features' scores in practical applications. Thus, semi-supervised feature selection can play an important role.

In this paper, we propose a novel semi-supervised filter method called constraint compensated Laplacian score (CCLS), which is similar to Laplacian score. The difference is that CCLS uses constraint information generated from a small amount of labeled data to compensate for the construction of local structure and global structure instead of unsupervised construction. Hence, CCLS has better locality discrimination ability than LS.

The outline of this paper is as follows: the background and motivation of this paper are given in Section 2. Section 3 enumerates several main methods used in feature selection. The CCLS method is presented in Section 4. Section 5 depicts the experimental setup and analyzes the results. Finally, conclusions are given in Section 6.

2 Semi-supervised feature selection for audio classification

Audio segmentation is the task of splitting an audio stream into segments of homogeneous content. Given a predefined set of audio classes, the process of segmentation involves joint boundary detection and classification, resulting in identification of segment regions as well as classification of those regions. Assuming that an audio signal has been divided into a sequence of audio segments using fixed window segmentation, our works focus on categorizing these audio segments into a set of predefined audio classes. Although there may be some differences between the traditional definition of audio classification and that in our work, the essential issues are the same.

Figure 1 illustrates the process of audio classification. In an audio classification system, every audio signal is first divided into mid-length segments which range in duration from 0.5 to 10 s. After this, the selected features are extracted for each segment using short-term overlapping frames. The sequence of short-term features in each segment is used to compute feature statistics, which are used as inputs to the classifier. In the final classification stage, the classifier determines a segment-by-segment decision.

In audio analysis and classification, there are dozens of features which can be used. A number of novel feature extraction methods have been proposed in recent years [17–19]. In this paper, some classical and widely used acoustic features are selected for feature selection sources. Widely used time-domain features [5] include short-term energy [20], zero-crossing rate [21], and entropy of energy [22]. Common frequency-domain features include spectral centroid, spectral spread, spectral entropy [23], spectral flux, spectral roll-off, Mel-frequency cepstral coefficients (MFCCs), and chroma vector [24].

There is a lot of complementary information among these features which can improve classification accuracy when used together; however, there is also a lot of redundant and even contradictory information which can degrade performance. It is hard to judge which combination of features is most likely to have a positive effect on classification. Furthermore, it is computationally infeasible to select the optimal feature subset by

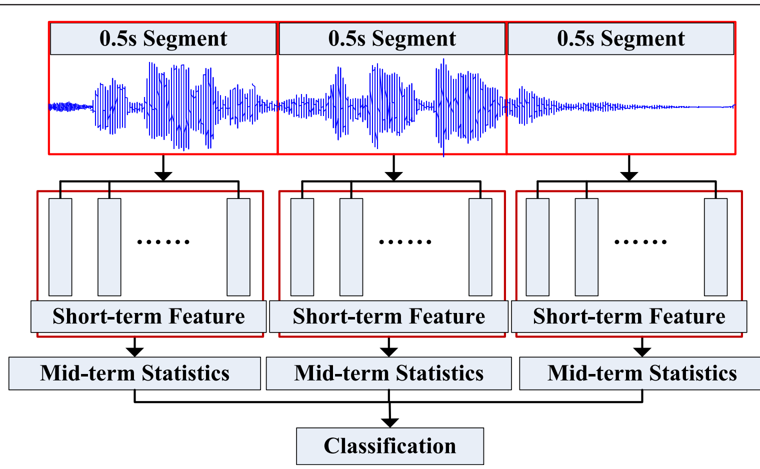

Fig. 1 The audio classification framework

exhaustive search. Thus, it is important to implement an effective feature selection method for this task.

Most supervised feature selection methods are dependent on labeled data. Unfortunately, it is difficult to obtain sufficient labeled data for audio classification, while unlabeled data is readily available. Semi-supervised feature selection methods can take good use of both labeled and unlabeled data; thus, this approach is more practical.

3 Related work

Let the training dataset with N instances be $X = \{\mathbf{x}_i \in \mathbb{R}^M\}$, where $i = 1, 2, \cdots, N$. Let $\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_M$ denote the corresponding feature vectors, where f_{ri} denotes the rth feature of $\mathbf{x}_i$, where $r = 1, 2, \cdots, M$. In semi-supervised learning, the training dataset X can be divided into two subsets. The first contains data $X^l = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_L\}$ with labels $Y^l = \{y_1, y_2, \cdots, y_L\}$, where $y_i = 1, 2, \cdots, C$ and C is the number of classes. The second set has only the unlabeled data $X^u = \{\mathbf{x}_{L+1}, \mathbf{x}_{L+2}, \cdots, \mathbf{x}_N\}$.

3.1 Laplacian score

Laplacian score is a recently proposed unsupervised feature selection method [8]. The basic idea is to evaluate features according to their locality preserving ability. If two data points are close to each other, they belong to the same class with high probability, so local structure is more important than global structure. The Laplacian score of the rth feature is a measure of local compactness computed as follows:

$$L_r = \frac{\sum_{i,j}\left(f_{ri}-f_{rj}\right)^2 S_{ij}}{\sum_i (f_{ri}-u_r)^2 D_{ii}}, \tag{1}$$

where $u_r = \sum_{i=1}^{N} f_{ri}/N$ denotes the mean of the rth feature of the whole data set. $\mathbf{D}$ is a diagonal matrix with $D_{ii} = \sum_j S_{ij}$, and $\mathbf{S}$ denotes the similarity matrix whose elements are defined as follows:

$$S_{ij} = \begin{cases} w_{ij} & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are neighbors} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The similarity w_{ij} between $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined by:

$$w_{ij} = e^{-\frac{\left\|x_i-x_j\right\|^2}{2\sigma^2}}, \tag{3}$$

where σ is a constant. $\mathbf{x}_i$ and $\mathbf{x}_j$ are considered to be neighbors if $\mathbf{x}_i$ is among the k nearest neighbors of $\mathbf{x}_j$ or $\mathbf{x}_j$ is among the k nearest neighbors of $\mathbf{x}_i$ in terms of Euclidean distance.

In the score function in Eq. 1, the numerator indicates the locality preserving the power of f_r, with smaller values indicating more local compactness in the feature space. The denominator is the weighted global variance of f_r. Thus, the criterion of the Laplacian score approach is to minimize the relative local compactness given by Eq. 1.

3.2 Constraint score

Constraint score is a supervised feature selection algorithm [9] which requires a relatively small amount of labeled data. For any pair of instances $(\mathbf{x}_i, \mathbf{x}_j)$ in the labeled data set X^l, there is a constraint assigned, either must-link (ML) or cannot-link (CL). The ML constraint is constructed if $\mathbf{x}_i$ and $\mathbf{x}_j$ have the same label, and the CL constraint is formed when $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to different classes. Then, ML and CL constraints are grouped into two sets Ω_{ML} and Ω_{CL}, respectively.

In the constraint score approach, the pairwise constraints between all pairs of data points are generated using the data labels, and a score function is computed as the following:

$$C_r = \frac{\sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\Omega_{\mathrm{ML}}}\left(f_{ri}-f_{rj}\right)^2}{\sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\Omega_{\mathrm{CL}}}\left(f_{ri}-f_{rj}\right)^2}. \tag{4}$$

This score represents a ratio of pairwise distances between same-class pairs and different-class pairs. Features are selected through minimizing this constraint score, with maximizes class separability.

3.3 Constrained Laplacian score

3.3.1 The score function

Constrained Laplacian score [10, 11] combines the above methods. The objective function of CLS is as follows:

$$\phi_r = \frac{\sum_{i,j}\left(f_{ri}-f_{rj}\right)^2 \mathcal{S}_{ij}}{\sum_{i,j}\left(f_{ri}-\alpha_{rj}^i\right)^2 \mathcal{D}_{ii}}, \tag{5}$$

where $\mathcal{S}_{ij} = S_{ij} + \mathcal{N}_{ij}$, with S_{ij} computed as in Eq. 2 from both labeled and unlabeled data and $\mathcal{N}_{ij}$ is given as follows:

$$\mathcal{N}_{ij} = \begin{cases} -w_{ij} & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are neighbors} \\ & \text{and } (\mathbf{x}_i, \mathbf{x}_j)\in\Omega_{\mathrm{ML}} \\ w_{ij}^2 & \text{if } [\mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are neighbors and } (\mathbf{x}_i, \mathbf{x}_j)\in\Omega_{\mathrm{CL}}] \text{ or} \\ & [\mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are not neighbors and } (\mathbf{x}_i, \mathbf{x}_j)\in\Omega_{\mathrm{ML}}] \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

In addition, $\mathcal{D}_{ii} = \sum_j \mathcal{S}ij$, and α_{rj}^i is defined as follows:

$$\alpha_{rj}^i = \begin{cases} f_{rj} & \text{if } (\mathbf{x}_i, \mathbf{x}_j)\in\Omega_{\mathrm{CL}} \\ u_r & \text{if } i=j \text{ and } x_i\in \mathrm{X}^u \\ f_{ri} & \text{otherwise.} \end{cases} \tag{7}$$

CLS combines Laplacian score to represent the internal structure characteristics of the entire data space and constraint score to incorporate class separability of the labeled data. However, this algorithm may be not suitable for some scenarios, as discussed in the next section.

3.3.2 The shortcomings of CLS

CLS uses constraint information from labeled data to help construct the local structure, represented by a matrix with elements N_{ij}. The elements of the similarity matrix used for local structure construction are $S_{ij} = S_{ij} + N_{ij}$, where S_{ij} is defined as follows:

$$\mathcal{S}_{ij} = \begin{cases} 0 & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are neighbors and } (\mathbf{x}_i, \mathbf{x}_j) \in \Omega'_{\mathrm{ML}} \\ w_{ij} + w_{ij}^2 & \text{if } [\mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are neighbors and } (\mathbf{x}_i, \mathbf{x}_j) \in \Omega'_{\mathrm{CL}}] \text{ or} \\ & [\mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are not neighbors and } (\mathbf{x}_i, \mathbf{x}_j) \in \Omega'_{\mathrm{ML}}] \\ w_{ij} & \text{if } [\mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are neighbors}] \text{ and } [\mathbf{x}_i \in \mathrm{X}^u \text{ or } \mathbf{x}_j \in \mathrm{X}^u] \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

As shown in Eq. 8, when two samples with the same labels are close to each other, the similarity between them is set to 0. In other words, CLS does not use close neighbors to construct the local structure. However, these example pairs are of high importance, because the local structures of neighbors are the most reliable. The preservation of such structure is an important measure of feature quality.

Moreover, when the constraint information from labeled data has conflicts with the local structure, CLS adds an additional item w_{ij}^2 to the similarity. This is problematic for several reasons, including because S_{ij} may be greater than 1 (for example, $w_{ij} = 0.9$, $S_{ij} = 1.71$). This conflict may appear in two cases, when two samples are close to each other but have different labels, or when two samples are far from each other but have the same label. In the first case, we would like to decrease the similarity because of the label differences, but the CLS formula instead increases the similarity with the added term. In the second case, we would like to increase the similarity, as the CLS formula does, but only to a limited degree because w_{ij} is close to 0 and thus w_{ij}^2 is very close to 0.

4 Constraint compensated Laplacian score

4.1 Score function

The main advantage of the Laplacian score approach is its locality-preserving ability. However, due to the lack of prior supervised information, the accuracy of this method is not high. The constraint score approach selects features based on a small amount of labeled data but ignores unlabeled data. CLS combines these approaches, but it neglects some important factors in estimations of local structures and supervised information.

To address these problems, we propose a new feature selection algorithm called constraint compensated Laplacian score (CCLS). The score function to be minimized is defined as follows:

$$\eta_r = \frac{\sum_{i,j} \left(f_{ri} - f_{rj} \right)^2 (S_{ij} + \bar{\mathcal{N}}_{ij})}{\Sigma_r + \Sigma_r^b - \Sigma_r^w}, \tag{9}$$

where

$$\bar{\mathcal{N}}_{ij} = \begin{cases} 1 - w_{ij} & \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are neighbors and } (\mathbf{x}_i, \mathbf{x}_j) \in \Omega_{\mathrm{ML}}. \\ -\gamma w_{ij} & \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are neighbors and } (\mathbf{x}_i, \mathbf{x}_j) \in \Omega_{\mathrm{CL}}. \\ \lambda & \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are not neighbors and } (\mathbf{x}_i, \mathbf{x}_j) \in \Omega_{\mathrm{ML}}. \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

where γ and λ are required parameters to be determined. S_{ij} is computed as in Eq. 2 using both labeled and unlabeled data, and

$$\Sigma_r = \sum_i (f_{ri} - \mu_r)^2 D_{ii} \tag{11}$$

$$\Sigma_r^b = \sum_c n_c \left(u_r^{(c)} - \mu_r^l \right)^2 \tag{12}$$

$$\Sigma_r^w = \sum_c n_c \left(\sigma_r^{(c)} \right)^2, \tag{13}$$

where n_c is the number of instances of the cth class, $\mu_r^l = \sum_{i|\mathbf{x}_i \in \mathrm{X}^l} f_{ri}/L$ is the mean of the rth feature of the labeled dataset, $\mu_r^{(c)} = \sum_{i|y_i = c} f_{ri}/L_c$ and $\left(\sigma_r^{(c)} \right)^2$ denote the mean and variance of the rth feature of the cth class, and L_c is the number of instances which belong to the cth class in the labeled dataset X_l.

4.2 Benefits of the new approach

The proposed CCLS approach integrates the LS and CS techniques under a unified semi-supervised framework, with two additional improvements: more accurate estimation of local structure and variance.

Table 1 Statistics of the UCI data sets

Dataset	Size	*M*	*C*	*N*	*L*
Ionoshpere	351	34	2	176	20
Segment	2310	19	7	1155	350
Soybean	47	35	4	24	12
Vehicle	846	18	4	400	100

M number of potential features, *C* number of classes, *N* number of instances, *L* number of labeled instances

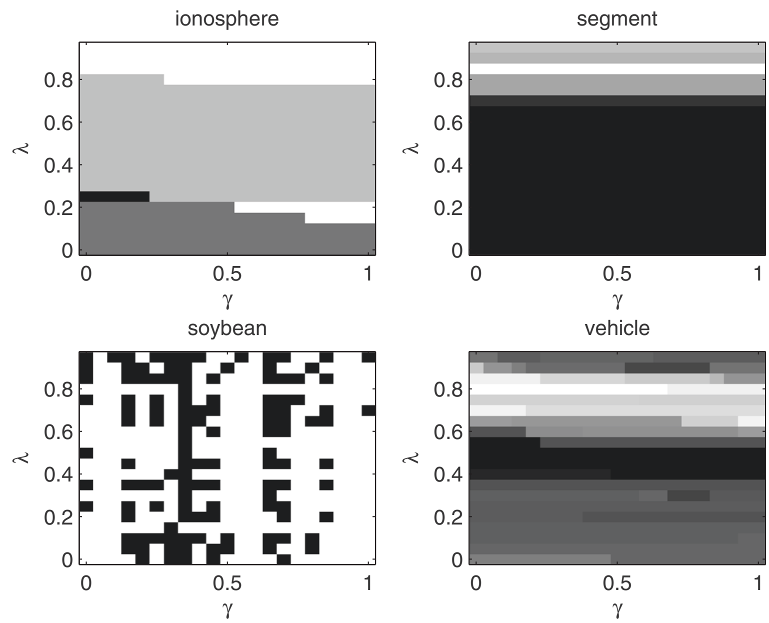

Fig. 2 Performance as a function of γ and λ. *Whiter shades* indicate better performance, *darker shades* indicate worse performance

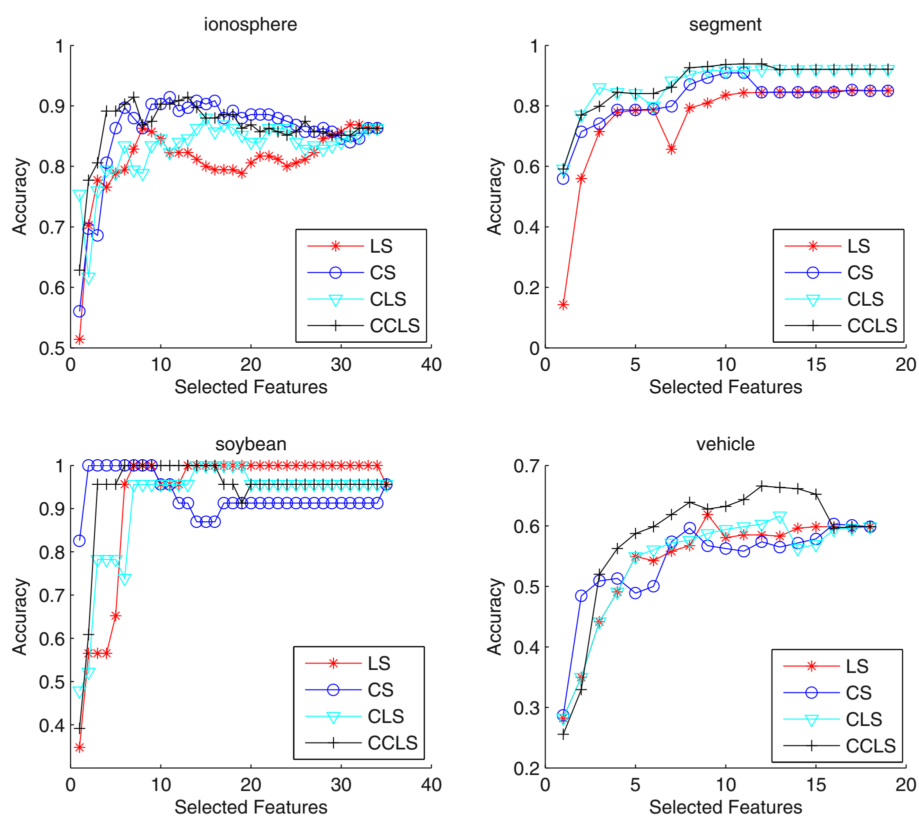

Fig. 3 Accuracy as a function of the number of selected features on four UCI data sets

4.2.1 The estimation of local structure

With respect to the calculation of within-class variance represented by the numerator of Eq. 9, the new CCLS method improves over CLS in the following aspects:

- When x_i and x_j are neighbors and also in the same labeled class, it is more certain that x_i is similar to x_j. It is more intuitive to increase the similarity between them, as represented in Eq. 10, rather than set it to zero as in Eq. 6 of CLS.
- When x_i and x_j are neighbors and in two different labeled classes, any local structure between x_i and x_j may mislead feature selection. Thus, it is appropriate to decrease the similarity instead of increasing them, as now represented in the second case of Eq. 9.
- When x_i and x_j are not neighbors but are in the same labeled class, they can be still considered as neighbors. In such most cases, however, the value of w_{ij} is very close to 0 because the distance between the points is large, so rather than using this distance as a weight, the new approach uses a controllable constant λ.

4.2.2 The estimation of variance

The new CCLS approach improves the accuracy of the variance of estimation. In the CLS approach, the variance estimation ignores the inner-class covariance of labeled data which has good discriminative ability. Moreover, CLS directly sums the variances of unlabeled data with that of labeled pairs from different classes. Specifically, in CLS, the variance of the rth feature vector $\mathbf{f}_r$ is given as follows:

$$\sum_{i,j}\left(f_{ri}-\alpha^i_{rj}\right)^2\mathcal{D}_{ii} = \sum_{i|\mathbf{x}_i\in X^u}\left(f_{ri}-\mu_r\right)^2\mathcal{D}_{ii}+\sum_{i,j|\left(\mathbf{x}_i,\mathbf{x}_j\right)\in\Omega_{CL}}\left(f_{ri}-f_{rj}\right)^2\mathcal{D}_{ii}. \tag{14}$$

In the proposed CCLS approach, the denominator of Eq. 9 shows that both inter-class covariance and inner-class covariance are used to estimate variance. This approach is motivated by the discrimination of these two types of covariance given by linear discriminative analysis

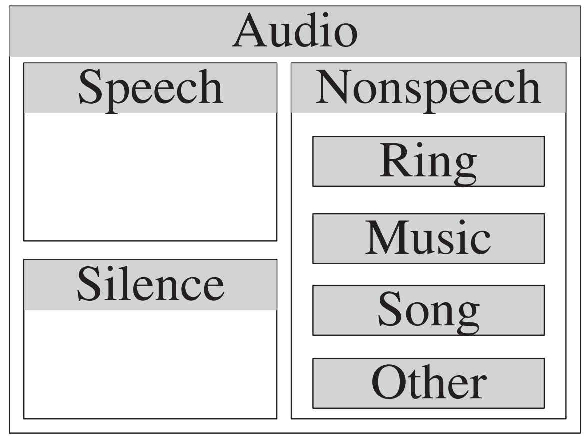

Fig. 4 The audio classes in telephone channel

[25]. Thus, a relevant feature should be correlated not only with larger variance of unlabeled data but also with larger inter-class covariance and smaller inner-class covariance.

4.3 Comparison of LS, CS, and CCLS

To illustrate the performance of LS, CS, CLS, and the proposed CCLS algorithm, we compare these four algorithms on several high-dimensional machine learning databases [26] including Ionoshpere, Image Segmentation, Soybean, and Vehicle datasets. The data set information and the sizes of whole training data set and labeled data set are detailed in Table 1. A nearest neighbor (1-NN) classifier with Euclidean distance is employed for classification.

To determine the parameters γ and λ, which govern how the rules affect feature selection performance, we experimentally vary parameter pairs from 0 to 1 with 0.05 intervals. The results are shown in Fig. 2, with lighter shades indicating better performance. Although the pattern is inconsistent, performance is generally

Table 2 Average accuracy of four different algorithms on UCI data sets

Algorithms	LS	CS	CLS	CCLS
Ionoshpere	80.87 ± 6.26	85.60 ± 7.22	83.55 ± 5.02	**86.35 ± 4.66**
Segment	76.26 ± 16.9	81.75 ± 8.10	75.31 ± 8.48	**82.29 ± 8.48**
Soybean	89.32 ± 15.7	93.04 ± 4.50	83.11 ± 19.2	**94.29 ± 11.5**
Vehicle	54.02 ± 9.27	54.07 ± 7.45	55.80 ± 8.82	**58.07 ± 11.2**

Best results of each experiment are set in bold

Table 3 Individual classification accuracy of different features

Feature	Dimensional	Accuracy		
		Mean	STD	Mean and STD
Zero-crossing rate	1	73.73	74.86	75.10
Short-term energy	1	45.81	46.03	69.41
Energy entropy	1	71.86	69.10	74.99
Spectral centroid	2	79.19	74.49	84.79
Spectral entropy	1	69.33	74.27	76.86
Spectral flux	1	79.21	69.09	77.86
Spectral roll-off	1	71.80	74.20	74.13
MFCCs	13	**84.26**	**86.44**	**87.66**
Harmonic	2	69.99	82.90	83.13
Chroma vector	12	83.49	83.87	83.73
All	35	81.07	85.97	86.04

Best results of each experiment are set in bold

Table 4 Averaged accuracy of different algorithms (400 labeled segments)

Alg.	Spec	ReliefF	LS	CS	CLS	CCLS
Ave.	85.26 ± 4.66	86.41 ± 2.79	85.32 ± 3.07	84.62 ± 2.92	83.40 ± 4.56	**88.46 ± 3.67**
Opt.	89.97	90.90	89.08	88.95	88.27	**91.14**
Num.	23	**19**	33	39	47	26

Best results of each experiment are set in bold

better when γ is high and λ is near 0.5. Thus, the parameters γ and λ are set to be 0.9 and 0.5, respectively, in all of our experiments followed.

The experimental results on UCI data sets are shown in Fig. 3 and Table 2. Figure 3 shows the plots for accuracy versus the number of selected features, and Table 2 compares the averaged accuracy across these cases. It can be seen that the performance of CS and CCLS is better than that of LS in all cases. This illustrates that constraint information from labeled data and local geometrical structure from unlabeled data are complementary, and using them in conjunction can be useful for feature selection.

5 Experiments and results

To further illustrate the effectiveness of CCLS, it is compared to several established feature selection methods. These include spectral feature selection (Spec) [27], ReliefF [28], Laplacian score, constraint score, and constrained Laplacian score.

5.1 Data and experimental setup

Experiments were performed using audio signals under telephone channel. Thus, each audio segment may contain speech, non-speech, or silence, with more detailed classes as shown in Fig. 4. “Speech” indicates direct dialogues between the calling and called users, when the call is connected, while “silence” implies the segment with comfort noise. “Non-speech” can be sub-classified into four types: ring, music, song, and other. “Ring” contains the single-tone, dual-tone, or multi-tone used for dialing or waiting warning, “music” and “song” are the waiting music before the call is connected or the environmental noise when the phone is in call. “Others” includes special sounds, such as laugh, barking, coughing, or other isolated sounds. Mixed segments, such as speech over music, are excluded from the dataset.

The database used here has been collected and manually labeled by Tsinghua University. It contains about 7 h of audio from 837 real telephone recordings. The speaker in each recording is different, as is the background music. The corpus consists of 3.4 h of speech data, 0.2 h of ring data, 0.1 h of music data, 0.1 h of song data, and 0.02 h of other data.

According to the label, an audio signal, which contains speech or non-speech, is divided into several 0.5-s segments. For each segment, all features mentioned in Section 2 are extracted based on the short-term analysis,

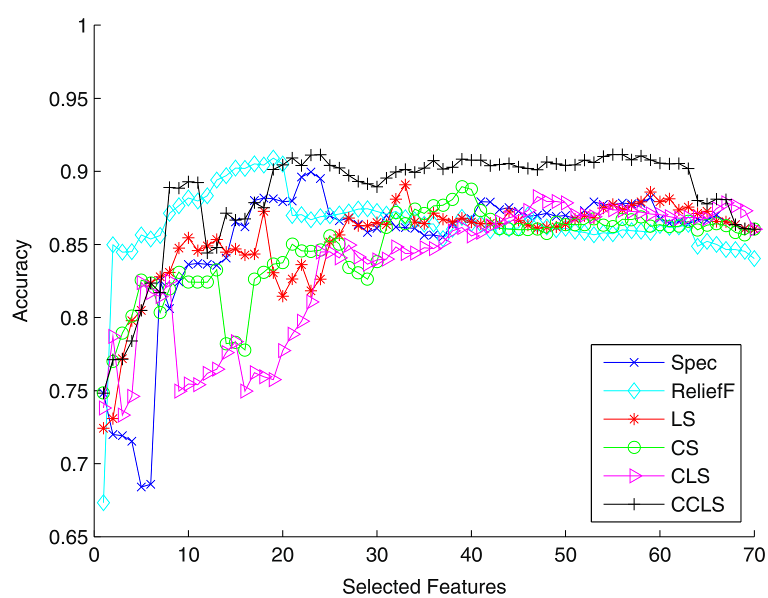

Fig. 5 Accuracy as a function of the number of selected features

and the dimension of short-term feature is 35. The frame length and frame step size are 32 and 10 ms, respectively. Then, the two mid-term statistics, mean value, and standard deviation are drawn per feature, so that the dimension of mid-term statistics vector is 70.

For feature selection, we choose 2000 speech segments and 2000 non-speech segments, with only 400 randomly chosen labeled segments. The γ value is set to 0.9 and λ = 0.5. We compare CCLS with unsupervised Laplacian score, as well as supervised constraint score, constrained Laplacian score, Spec, and ReliefF. We use a development dataset containing 200 speech segments and 200 non-speech segments to choose the optimal feature subset. The test dataset includes 500 speech segments and 500 non-speech segments.

In all experiments, the k-nearest neighborhood (KNN) classifier with Euclidean distance and $k = 5$ is utilized for classification after feature selection. To avoid the influence of the classifier, the training datasets of the classifier for all experiments are kept the same.

5.2 Experimental results

Ten types of short-term features extracted are listed in Table 3. Two statistics, mean and standard deviation are used as the mid-term representation of the audio segments. Table 3 shows the classification accuracy of different features for audio classification. The top three best features are MFCCs, chroma vector, and spectral centroid, and the worst feature is short-term energy. Moreover, using all of these features does not improve but rather decreases the accuracy, as seen by comparing MFCC accuracy to that using all features, indicating that there is redundant and even contradictory information among the features. Thus, it is valuable to use feature selection as a preprocessing module.

Table 4 compares the averaged accuracy (Ave.), optimized accuracy (Opt.), and the optimized number (Num.) of features among all evaluated methods, and the value after the symbol "±" denotes the standard deviation. Results indicate that the performance is significantly improved by using the first d features selected from the ranking list of features generated by feature selection algorithms. This supports the hypothesis that there is redundant and even contradictory information among the original feature space and that a feature selection algorithm can remove irrelevant and redundant features effectively.

CCLS is superior to other evaluated methods not only in terms of averaged accuracy but also in terms of optimized accuracy. In contrast, CLS has the lowest averaged accuracy and optimized accuracy. This is because the estimations of local structure and variance are not accurate for CLS method as described in Section 4.2.

Figure 5 shows accuracy vs. the number of selected features. It can be seen that the performance of CCLS is significantly better than that of Spec, Laplacian score, constraint score, and constrained Laplacian score. This supports that combining supervised information with data structures to evaluate the relevance of features is useful in feature selection.

Table 5 Performance of supervised and semi-supervised methods with 200 labeled segments

Algorithms	Spec	ReliefF	CS	CCLS
Ave.	83.33 ± 3.94	85.53 ± 5.92	81.06 ± 3.34	**87.62 ± 3.50**
Opt.	87.20	89.24	87.68	**91.64**
Num.	57	**35**	55	40

Best results of each experiment are set in bold

To explore the influence of the numbers of labeled segments on the performance of the algorithm, different numbers of labeled data are used. The averaged accuracy, optimized accuracy, and the optimized number of features on the condition of 200 and 800 labeled segments are summarized in Tables 5 and 6, respectively. Comparing Table 4 with Tables 5 and 6, it is easy to conclude that the performance improves as the number of labeled data segments increases from 200 to 800. The CCLS is best in terms of averaged accuracy and optimized accuracy regardless of the number of labeled segments. The performance of ReliefF is always better than others in terms of optimized number of features.

Figure 6 shows the plots of accuracy vs. the number of selected features and the amount of labeled data. However, it should also be noticed that the performances of CCLS and ReliefF do not drop rapidly when decreasing the amount of labeled data to 200, while the CS and Spec algorithms are unable to select relevant features.

In all cases, there are many irrelevant features, almost two-thirds, which can be removed to achieve the best performance. This not only improves classification accuracy but also reduces the time complexity of classification.

Figure 7 shows the plots of average accuracy vs. the number of labeled data segments. The average accuracy increases with the addition of labeled data, asymptoting between 500 and 700 segments. The CCLS algorithm outperforms the other algorithms significantly.

After the optimal feature subset has been selected, classification is done on the test data set. The results are listed in Table 7. The optimal feature subset selected

Table 6 Performances of supervised and semi-supervised methods with 800 labeled segments

Algorithms	Spec	ReliefF	CS	CCLS
Ave.	86.76 ± 2.82	88.47 ± 4.69	87.24 ± 3.48	**89.72 ± 3.71**
Opt.	89.49	92.27	91.45	**92.71**
Num.	55	**18**	23	25

Best results of each experiment are set in bold

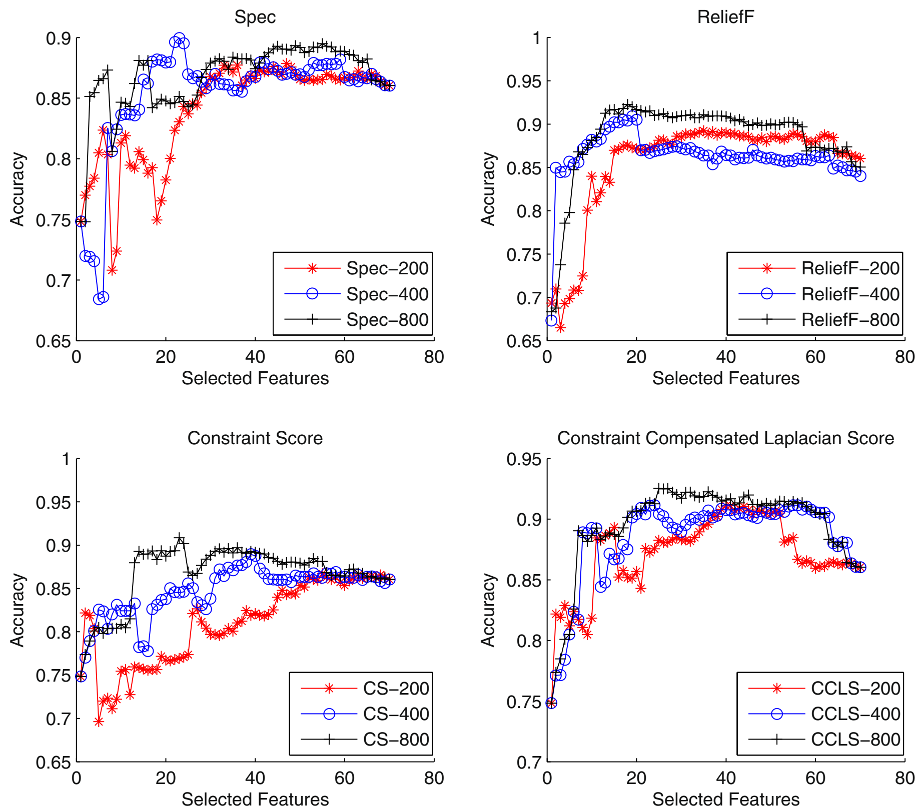

Fig. 6 Accuracy as a function of number of selected features and number of labeled data segments

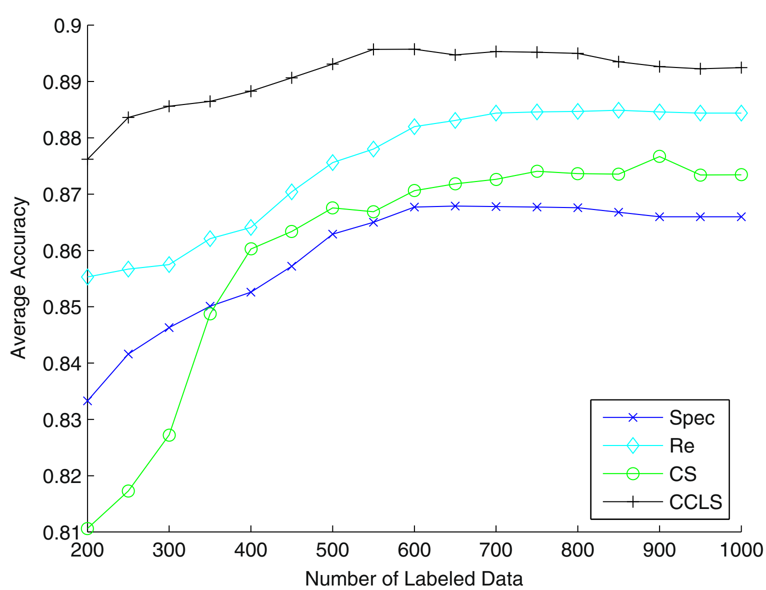

Fig. 7 Average accuracy as a function of the number of labeled data segments

Table 7 Accuracy of different algorithms on test dataset, using feature subsets tuned over the development dataset (400 labeled segments)

Algorithms	Spec	ReliefF	LS	CS	CLS	CCLS
Accuracy	88.87	90.24	89.77	89.02	87.35	**91.06**

Best results of each experiment are set in bold

from development dataset improves the performance on the test dataset. Though CCLS still outperforms other algorithms, the accuracy differences between algorithms are relatively small. However, the average accuracy of CCLS across all feature subsets is much high than the other algorithms, which indicates that the algorithm is more robust to feature subset selection than the comparative methods.

6 Conclusions

In this paper, we have presented a semi-supervised filter-based feature selection method. The new CCLS method integrates locality preservation across unlabeled data and label consistency within labeled data. Experimental results show that the proposed algorithm outperforms Spec, ReliefF, LS, and CS for audio classification.

As mentioned in Section 5.2, the performance of CCLS was not as good as that of ReliffF in terms of optimized number of features. This may indicate that there are some redundancy features in the optimum feature set selected by CCLS method. Several studies have addressed influences of such redundancy [11, 29, 30]. To improve the generalization quality of CCLS, there are primarily two areas for the future work: (1) Improvement of discrimination ability across audio classes, for example, more accurate estimation of local structure and variance, and (2) redundancy should be further removed from CCLS optimal feature sets.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (No. 61175017, No. 61403415, No. 61370034, and No. 61403224).

Author details

[1]Zhengzhou Information Science and Technology Institute, Zhengzhou, China. [2]Department of Electronic Engineering, Tsinghua University, Beijing, China. [3]Department of Electrical and Computer Engineering, Marquette University, Milwaukee, WI, USA.

Received: 20 August 2015 Accepted: 24 February 2016
Published online: 15 March 2016

References

1. S Zahid, F Hussain, M Rashid, MH Yousaf, HA Habib, Optimized audio classification and segmentation algorithm by using ensemble methods. Math. Probl. Eng. **2015**, 11 (2015). doi:10.1155/2015/209814. Article ID 209814
2. T Hirvonen, *Speech/Music Classification of Short Audio Segments* (Proc. 2014 IEEE International Symposium on Multimedia, Taichung, 2014). Dec. 10–12
3. Y Vaizman, B McFee, G Lanckriet, Codebook-based audio feature representation for music information retrieval. IEEE/ACM Transactions on Audio, Speech, and Language Processing **22**(10), 1483–1493 (2014)
4. P Mahana, G Singh, Comparative analysis of machine learning algorithms for audio signals classification. International Journal of Computer Science and Network Security **15**(6), 49–55 (2015)
5. Giannakopoulos, T, & Pikrakis, A (2014). *Introduction to Audio Analysis: A MATLAB Approach*. Elsevier Academic Press.
6. Z Zhao, H Liu, *Spectral Feature Selection for Data Mining (Data Mining and Knowledge Discovery Series)*(Chapman and Hall-CRC, Boca Raton, FL, USA, 2012)
7. H Yan, J Yang, Locality preserving score for joint feature weights learning. Neural Netw. **69**, 126–134 (2015)
8. X He, D Cai, P Niyogi, *Laplacian Score for Feature Selection* (in Proc NIPS, Vancouver, BC, Canada, 2005)
9. D Zhang, S Chen, Z Zhou, Constraint score: a new filter method for feature selection with pairwise constraints. Pattern Recogn. **41**(5), 1440–1451 (2008)
10. K Benabdeslem, M Hindawi, *Constrained Laplacian Score for Semi-Supervised Feature Selection* (in Proc ECML-PKDD, Athens, Greece, 2011), pp. 204–218
11. K Benabdeslem, M Hindawi, Efficient semi-supervised feature selection: constraint, relevance and redundancy. IEEE Trans. Knowl. Data Eng. **26**(5), 1131–1143 (2014)
12. B Efron, T Hastie, I Johnstone, R Tibshirani, Least angle regression. Ann. Stat. **25**, 407–449 (2004)
13. Mckinney, M, & Breebaart, J (2003). Features for Audio and Music Classification. *Proc. the International Symposium on Music Information Retrieval* (pp. 151-158). Baltimore, USA: Library of Congress.
14. Hao Jiang, Lie Lu, and HongJiang Zhang, A robust audio classification and segmentation method, Microsoft Research, No. MSR-TR-2001-79.
15. R Fiebrink, I Fujinaga, *Feature Selection Pitfalls and Music Classification* (in proc International Conference on Music Information Retrieval, Victoria, Canada, 2006)
16. Maria Markaki (2011). *Selection of Relevant Features for Audio Classification tasks*. Doctoral Dissertation, Department of Computer Science, University of Crete.
17. JT Geiger, B Schuller, G Rigoll, *Large-Scale Audio Feature Extraction and SVM for Acoustic Scene Classification* (in Proc Applications of Signal Processing to Audio and Acoustics, New Paltz, 2013), pp. 1–4
18. T Ramalingam, P Dhanalakshmi, Speech/music classification using wavelet based feature extraction techniques. J. Comput. Sci. **10**(1), 34–44 (2014)
19. S Zubair, F Yan, W Wang, Dictionary learning based sparse coefficients for audio classification with max and average pooling. Digital Signal Processing **23**(5), 960–970 (2013)
20. C Panagiotakis, G Tziritas, A speech/music discriminator based on RMS and zero-crossings. IEEE Trans. Multimedia **7**(1), 155–166 (2005)
21. E Scheirer, M Slaney, *Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator* (in Proc. ICASSP, Munich, Germany, 1997)
22. Giannakopoulos, T, Pikrakis, A, Theodoridis, S (2008). Gunshot detection in audio streams from movies by means of dynamic programming and Bayesian networks. *Proc. of ICASSP* (pp. 21-24). Las Vegas, USA: IEEE
23. H Misra, S Ikbal, H Bourlard, and H HERMANSKY, "Spectral entropy based feature for robust ASR, in Proc. ICASSP, 2004.
24. MA Bartsch, GH Wakefield, Audio thumbnailing of popular music using chroma-based representations. IEEE Trans. Multimedia **7**(1), 96–104 (2005)
25. K Fikunaga, *Introduction to Statistical Pattern Recognition*, 2nd edn. (Academic, San Diego, 1990)
26. C Blake, E Keogh, CJ Merz, *UCI Repository of Machine Learning Databases* (Department of Information and Computer Science, University of California, Irvine, 1998). http://www.ics.uci.edu/~mlearn/MLRepository.html
27. Zhao, Z, & Liu, Hin (2007). Spectral Feature Selection for Supervised and Unsupervised Learning. *Proc. ICML* (pp. 1151-1157). New York, USA: ACM
28. M Robnik-Šikonja, I Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF. Mach. Learn. **53**, 23–69 (2003)
29. C Ding, HC Peng, *Minimum Redundancy Feature Selection from Microarray Gene Expression Data* (in Proc. IEEE CSB, Stanford, CA, 2003), pp. 523–528
30. L Yu, H Liu, Efficient feature selection via analysis of relevance and redundancy. J. Mach. Learn. **5**, 1205–1224 (2004)