Distributed Sparse Approximation for Frog Sound Classification

Bo Wei[†‡], Mingrui Yang[‡], Rajib Kumar Rana[‡], Chun Tung Chou[†], and Wen Hu[‡]
[†]School of Computer Science and Engineering, University of New South Wales, Sydney, Australia
[‡]Autonomous Systems Laboratory, CSIRO ICT Centre, Australia
{bwei, ctchou}@cse.unsw.edu.au, {mingrui.yang, rajib.rana, wen.hu}@csiro.au

ABSTRACT

Sparse approximation has now become a buzzword for classification in numerous research domains. We propose a distributed sparse approximation method based on ℓ_1 minimization for frog sound classification, which is tailored to the resource constrained wireless sensor networks. Our pilot study demonstrates that ℓ_1 minimization can run on wireless sensor nodes producing satisfactory classification accuracy.

Categories and Subject Descriptors

C.2.1 [**Computer Communication Networks**]: [Network Architecture Design]

General Terms

Algorithms, Design, Experimentation, Performance

Keywords

ℓ_1 minimization, sparse approximation

1. INTRODUCTION

Sparse approximation, more precisely ℓ_1 minimization has recently been adopted largely by the researchers in the image processing and computer vision domain for efficient classification. In particular, researchers from the field of face classification [2] have reported that ℓ_1 minimization offers better classification accuracy compared to the state of the art classification algorithms such as Support Vector Classification. Motivated by this promising outcome we are interested in investigating the performance of ℓ_1 minimization for frog sound classification within a wireless sensor network platform.

There are two key challenges that need to be addressed before ℓ_1 minimization can be used for frog sound classification in wireless sensor networks. First, scaling down the minimization problem so that it can be solved on a wireless sensor network node. Second, unlike the test samples in the face or facial expression application, test samples may consist of multiple frog sounds. Therefore, ℓ_1 minimization is expected to classify multiple classes simultaneously. In our work, we seek to address these two key challenges.

IPSN'12, April 16–20, 2012, Beijing, China.
ACM 978-1-4503-1227-1/12/04.

2. CLASSIFICATION BY ℓ_1 MINIMIZATION

In this paper we propose a system for conducting frog sound classification using wireless sensor networks. In the proposed system, collaborative or cooperative classification is conducted locally at the sensor nodes. Upon detecting a frog sound, a sensor node organizes a cluster including other wireless sensor nodes and decides a cluster head. The cluster head distributes the classification task among the cluster nodes. Each cluster node conducts a part of the classification tasks and transmits the classification result to the cluster head. The cluster head joins the individual classification results to a complete result and identify the frog class(es).

Before describing our proposed method for frog sound classification, let us describe the general model for classification using ℓ_1 minimization. Consider we have c training classes and each class has t number of training subjects or samples. Let us define a matrix A containing the entire training set. Therefore $A = [a_{1,1}, a_{2,1}, \dots, a_{t,c}] \in \mathbb{R}^{m\times(t\times c)}$. A test subject or sample $y \in \mathbb{R}^m$ belonging to the ith class can be represented as a linear combination of the training samples:

$$y = A\alpha \in \mathbb{R}^m, \tag{1}$$

where $\alpha = [0, 0, \dots, \alpha_{1,i}, \alpha_{2,i}, \dots, \alpha_{t,i}, 0, 0, \dots]^T$, $\alpha_{j,i} \in \mathbb{R}$, is a coefficient vector whose elements are zeros except for the ith training class.

Since it is now well known that ℓ_1 minimization can be used to find the sparse solution to Eq. (1) and the solution can be calculated in polynomial time, [2] used this idea for face recognition. However, notice that classification is not based on the largest sparse coefficients, but is given by

$$\arg\min_i r_i(y) = \|y - A\delta_i(\alpha)\|_2, \tag{2}$$

where $\delta_i(\alpha)$ selects only the nonzero coefficients belonging to class i.

In our frog song classification problem, audio stream of a given class of frog is segmented into smaller windows and the Fourier transform of each window ($a_{i,j}$) of the sound is used as a column of the training matrix A. A test frog sound is first segmented by the window and then transformed to the frequency domain. Then the residual with respect to each class is calculated using Eq. (2) and the class with the minimum residual is identified as the match.

Notice that the dimension of the training matrix is typically very large. Finding the solution to Eq. (1) requires to run an ℓ_1 minimization problem, which is already computationally expensive for wireless sensor nodes. Therefore, in order to use the ℓ_1 based classification on wireless sensor nodes we propose to segment the training matrix into smaller parts where each part forms an Eq. (1) and is solved in a separate sensor node. For example, if p nodes take part in the classification, the training matrix A is divided into p parts as $[A_1, A_2, \ldots, A_p]$ and the observation made by the initiating node is transmitted to all the participating nodes. Then each node solves for α

$$y = A_i \alpha_i, \quad i = 1, \ldots, p.$$

$\{\alpha_i\}_{i=1}^{p}$ are then transmitted back to the initiating node, which forms the coefficient vector $\alpha = [\alpha_1^T, \alpha_2^T, \ldots, \alpha_p^T]^T$, and calculates the residuals using Eq. (2) to identify the frog sound.

3. RESULTS

We perform the frog sound classification in Matlab R2010b. Gradient Projection for Sparse Reconstruction (GPSR) (see [1]) is applied to calculate the coefficients for the training set.

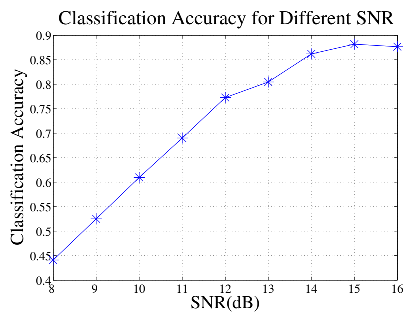

Figure 1: Accuracy Rate for Different SNR

In the simulation environment, the sample rate is 15,000Hz, and the window size is 0.5 seconds. Fig. 1 shows the accuracy rate of our ℓ_1 minimization approach using GPSR with respect to different signal to noise ratio (SNR). The noise is collected from the frog's living environment. We vary its amplitude and add it to the original test sound to get different SNR. The test sound sample set has 14 species whose call duration varies from 7.0 seconds to 43.4 seconds. We segment the test sample according to the window size and count the correct number of windows of the test sound, and compare it to the total number of windows to get the accuracy. As we expect, the accuracy rate increases to near 90% as SNR increases from 8 to 16.

Next, we show some preliminary results of the distributed frog sound classification. We want to classify the testing sound from Cyclorana cryptotis (class 2). Fig. 2 shows the classification residual by using one node, two nodes, and three nodes respectively. As we can see from these plots, Cyclorana cryptotis always has the smallest residual among all training species. The main concern here is how far we can go. It is clear that we can not go through this process forever, based on both the resource limitation and the accuracy concern. We are still working on finding a critical point to balance the trade off between efficiency and accuracy.

We also tested the performance of our approach on multi-frog detection. Encouragingly, when multi-class frog sounds appear simultaneously in the test sample, we are still able to detect these classes using this approach. As we can see from

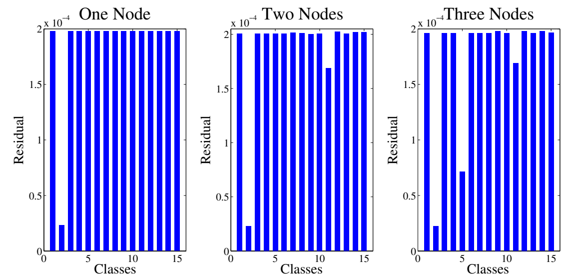

Figure 2: Residual Plots for Multi-node Classification

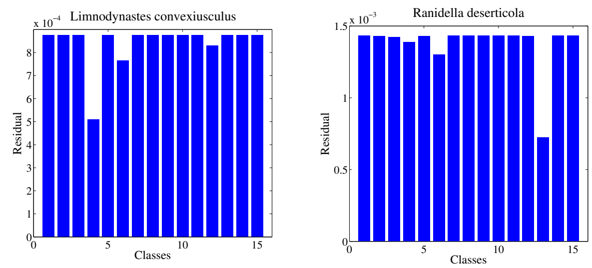

Figure 3: Residual Plots for Multi-frog Detection

the residual plots, Fig. 3, when Limnodynastes convexiusculus and Ranidella deserticola are singing simultaneously, we are still able to classify them using ℓ_1 minimization and Eq. (2). Therefore, ℓ_1 minimization also makes the simultaneous classification for multi-class test samples feasible.

4. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel distributed frog sound classification method based on ℓ_1 minimization. We scale down the typical ℓ_1 base classification problem to be solvable on resource improvised wireless sensor nodes. We also demonstrate that ℓ_1 minimization is feasible to classify multiple classes simultaneously. To the best of our knowledge we are the first to propose and evaluate the above two features of ℓ_1 based classification.

The key issue that we are to address in the full bloomed version of our work is, how to define the number of segments of the training matrix. Our simulation shows that increasing the number of segments improves the computation time. However, after certain number of segments, the classification accuracy diminishes. We seek to find the optimal number of segments that offers a good trade-off of classification accuracy and computation time.

5. REFERENCES

[1] M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *Selected Topics in Signal Processing, IEEE Journal of*, 1(4):587–596, 2008.

[2] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210–227, 2009.