A Context-Aware Method Based Cattle Vocal Classification for Livestock Monitoring in Smart Farm

Dr. Farook Sattar

Dept. of Electrical and Computer Engineering,
University of Victoria, BC, Canada; fsattar@ieee.org

1st International Online Conference on Agriculture - Advances in Agricultural Science and Technology (IOCAG2022), 10-25 Feb. 2022

Outline

- Introduction
- Data
- Method
- Results & Performance
- Conclusion

Introduction

As farming systems become increasingly automated, it is possible to dynamically adjust the environment in which the animals are kept and automatically change the temperature, lighting and ventilation.

With the help of sensor and artificial intelligence (AI) technologies, the farmers and farm owners can also detect diseases in animals and take immediate actions accordingly.

Introduction

The implementation of smart technologies in livestock farming helps in gathering and processing real-time data related to animals health and general behavior, including their feeding behavior, food and water quality, hygiene levels, and others.

For example, growing population of cattle with increasing dairy farms and increasing adoption of livestock monitoring technology in developing countries create a strong demand for livestock monitoring in smart farms.

Introduction

Here we aim to develop an automated acoustic analysis tool for livestock monitoring using contextual acoustic features and multiclass support vector machine (MSVM) classifier.

The proposed scheme can be used to assist welfare, production, and disease detection for farm animals from their vocalizations.

This would thereby enhance the smart farming/precision livestock farming to increase the agricultural production.

Data

We used an open access dataset[1] containing 270 cattle classification records collected from 12 recording sensors (USB mic, Shenzhen kobeton technology, Shenzhen, China, frequency response: 16 Hz-100 kHz, sensitivity: -47 dB ± 4 dB). The audio data are collected in three separate zone with 4 microphones are placed in each zone and located at a height of 3 m in three separate livestock facilities(see[2] for more details).

[1] https://www.mdpi.com/2076-2615/11/2/357/s1

[2] D.-H. Jung, N.Y. Kim, *et. al.*, "Deep learning-based cattle vocal classification model and real-time livestock monitoring system with noise filtering", *Animals*, vol 11, 2021, pp. 357.
doi: https://doi.org/10.3390/ani11020357.

Method - Outline

The basic idea of the proposed approach lies in the integration of auditory processing model and contextual information for extracting useful features. The method adopts the multiresolution framework. The general outline of the multiclass classification considered here for identifying cattle vocals, is shown below:

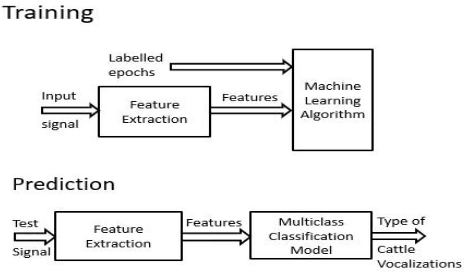

Figure: The general outline of multiclass classification for identifying cattle vocals.

Method - Data Preprocessing

The dataset of each vocalization is resampled (from 44,100 Hz to 16,000 Hz) and resized into N-sample data blocks ($N = 8192$ samples here, referring to 0.512 s) followed by time windowing using N-sample Hamming windows.

Note that resampling is done here to reduce the computational complexity, while resizing is performed to save memory by compressing the signal without changing the spectral content.

Method - Contextual Acoustic Features

- We have introduced a contextual acoustic features, which encodes the multi-resolution energy distributions in the time-frequency plan based on the cochleagram representation of an input signal.

- We incorporate a number of cochleagrams at different resolutions to design the new features set.

- The cochleagram with high resolution captures the local information, while the other low resolution cochleagrams capture the contextual information at different scales.

Method - Contextual Acoustic Features

To compute the cochleagram, we first pass an input signal to a gammatone filter bank, where the impulse response of a particular gammatone filter is given by

$$\begin{aligned} h(t) &= t^{(\eta-1)} e^{-2\pi B_{f_c} t} \cos(2\pi f_c t) \quad (t \geq 0) \\ &= 0 \quad (t \leq 0) \end{aligned} \tag{1}$$

Here in Eq.(1), the parameter η is the order of the filter, f_c denotes the center frequency while B_{f_c} refers to the bandwidth given f_c.

Method - Contextual Acoustic Features

The gammatone filter function is used in models of the auditory periphery representing critical-band filters where the center frequencies f_c are uniformly spaced on the equivalent rectangular bandwidth (ERB) scale.

The relation between B_{f_c} and f_c is given by

$$B_{f_c} = 1.019 \times ERB(f_c) = 1.019 \times 24.7(4.37 \times f_c/1000 + 1) \qquad (2)$$

Method - Contextual Acoustic Features

Then each output signal from the gammatone filter bank is divided into 20 ms frames with a 10 ms frame shift; the cochleagram is then obtained by calculating the energy of each time frame at each frequency channel.

Each T-F unit in the cochleagram contains only local information, which may not be sufficient to accommodate the diversity in the real-recorded input data.

Method - Contextual Acoustic Features

To compensate for this, a new feature set is designed providing contextual information by including the energy distribution in the neighborhood of each T-F unit. The steps for computing contextual acoustic features are as follows.

(1) Given input ocean data, compute the first 32-channel cochleagram (CB1) followed by a log operation applied to each T-F unit.

(2) Similarly, the second cochleagram (CB2) is computed with the frame length of 200 msec and frame shift of 10 msec.

(3) The third cochleagram (CB3) is derived by averaging CB1 using a rectangular window of size (5×5) including 5 frequency channels and 5 time frames centered at a given T-F unit. If the window goes beyond the given cochleagram, the outside units take the value of zero (i.e. zero padding).

Method - Contextual Acoustic Features

(4) The fourth cochleagram CB4 is computed in a similar way to CB3, except that a rectangular window of size (11×11) is used.

(5) Concatenate CB1-CB4 to generate a feature matrix F and integrate it along the time frame to obtain a set of contextual acoustic features of dimension (128×1).

Method - MSVM Classification

Separating various noisy cattle calls, is a multiple classification based monitoring problem, which is solved here by considering one-against-all optimization formulation based on Crammer and Singer (CS) method for a multiclass support vector machine (MSVM) providing fast convergence and high accuracy.

Method - MSVM Classification

In general, a MSVM classifier solves a d-class classification problem by constructing decision functions of the form:

$$x \mapsto \arg \min_{c \in \{1,\ldots,d\}} \{ < w_c, \phi(x) > + b_c \} \tag{3}$$

given *i.i.d.* training data $((x_1, y_1), \ldots, (x_l, y_l)) \in (X \times \{1, \cdots, d\})^l$.

Method - MSVM Classification

Here, $\phi : X \rightarrow \mathcal{H}, \phi(x) = k(x, \cdot)$, is a feature map into a reproducing kernel Hilbert space $\mathcal{H}$ with corresponding kernel k, and $w_1 \cdots, w_d \in \mathcal{H}$ are class-wise weight vectors and $< \cdot >$ refers dot product. The CS method is usually only defined for hypotheses without bias terms, that is, for b_c=0. This CS based MSVM classifier is trained by solving the primal problem

$$\min_{w_c} \frac{1}{2} \sum_{c=1}^{d} < w_c, w_c > + C \sum_{n=1}^{l} \eta_n \tag{4}$$

subject to

$$\forall n \in \{1, \cdots, l\}, \forall c \in \{1, \cdots, d\} \backslash \{y_n\} :< w_{y_n} - w_c, \phi(x_n) > \geq 1 - \eta_n$$

and

$$\forall n \in \{1, \cdots, l\} : \eta_n \geq 0$$

where η refers to the ‘slack’ variable. For learning structured data, CS's method is usually the MSVM algorithm of choice taking all class relations into account at once to solve a single optimization problem with fewer slack variables.

Results and Performance - Dataset Used

Four types of cattle calls namely food anticipation calls, estrus calls, cough sounds, and normal calls which have been used are shown in the following Table.

Table: Number of various cattle vocalizations (calls) used

Type of vocalization	Number of vocalization
Food anticipation	100
Estrus	117
Cough	11
Normal	42

Results and Performance - Visualizations

Here, the spectrograms of different audio samples corresponding to food anticipation, estrus, cough sound, and normal vocals are displayed for illustration.

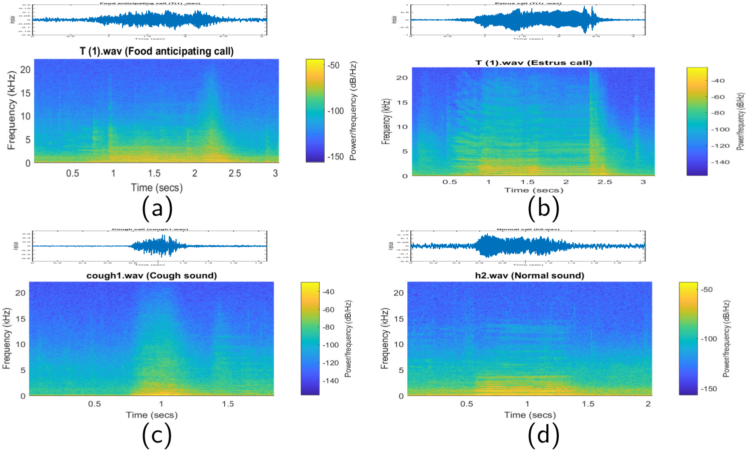

Figure: Illustrative spectrogram plots for various cattle vocal samples referring to (a) food anticipation call, (b) estrus call, (c) cough sound, (d) normal call.

Results and Performance - Contextual Features

In order to reduce the redundancy while maintaining the variability of the contextual features, the decimated version of the feature set is considered as a kind of feature selection. A decimation factor of 8 is used here which reduces the length of the features from 128 to 16. For illustration the average of all the (16×1) features for each types of cattle vocals are plotted in Figure 3.

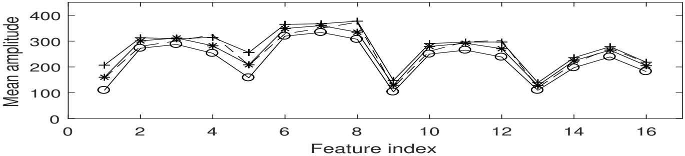

Figure: Plots for the average of the contextual features for each types of cattle calls; '-*-' : food anticipation, '-o-' : estrus, '-+-' : cough, '- -' : normal calls.

Results and Performance - Multiclass Classification

The proposed scheme is evaluated in terms of classification results for real recorded cattle calls. Results are obtained over 50 different runs in which the feature sets are split randomly by data samples where 70% of the data are used for training and 30% of the data are retained for testing/prediction.

In each case, the feature set is normalized to have zero mean and unit standard deviation. Here we have selected the default MSVM parameter C (regularization parameter) and γ (bandwidth parameter) of the radial Gaussian kernel $k(x; x^{'}) = \exp(-\gamma||x - x^{'}||^2)$ as C=10 and γ=2.

Results and Performance - Multiclass Classification

Table: Confusion matrix with the contextual feature set where the average classification accuracy (%) is shown in the right bottom corner (bold face) calculated from the confusion matrix as $\left(\frac{\text{Sum of diagonal elements}}{\text{Sum of all elements}}\right)$

	Food anticipation	*Estrus*	*Cough*	*Normal*	*Specificity*
Food anticipation	24	3	0	1	0.85
Estrus	6	29	0	0	0.82
Cough	0	0	1	1	0.50
Normal	1	0	0	9	0.90
Sensitivity	0.77	0.90	1	0.81	**84.00**

Results and Performance - Multiclass Classification

The average classification accuracy (%) for various feature size (M) are listed in Table 3, where $M = 16$ gives the best result by the proposed scheme.

Table: Average classification accuracy (%) for different feature size M

M	8	16	32
Average accuracy (%)	78.67	84.00	80.82

Results and Performance - Comparison Results

The comparison results with MFCC (mel frequency ceptral coefficients) features are presented in the following Table. The following parameters are set: MFCC window length=20 ms (320 samples), number of MFCC features=12, MFCC window overlapping=50% providing the best results with the MFCC features.

Results and Performance - Comparison Results

Table: Confusion matrix with the MFCC feature set where the average classification accuracy (%) is shown in the right bottom corner (bold face) calculated from the confusion matrix as $\left(\frac{\text{Sum of diagonal elements}}{\text{Sum of all elements}}\right)$

	Food anticipation	*Estrus*	*Cough*	*Normal*	*Specificity*
Food anticipation	23	4	1	0	0.82
Estrus	18	15	0	1	0.44
Cough	2	0	0	0	0
Normal	3	0	0	7	0.70
Sensitivity	0.50	0.78	0	0.87	**60.81**

Conclusion

- We introduce a new acoustical method for automatic livestock monitoring in smart farm.
- The proposed framework is found to be effective in classifying various types of cattle sounds analyzed herein.
- The performance of the proposed method is promising in terms of classification accuracy which outperforms the results obtained by the MFCC features.
- Future works include the use of larger dataset to improve the performance as well as analyze other types of vocalizations, e.g. poultry, sheep, with the aim to deliver the highest levels of animal welfare for precision livestock farming.