INTERACTIVE SINGING MELODY EXTRACTION BASED ON ACTIVE ADAPTATION

Kavya Ranjan Saxena Vipul Arora

Department of Electrical Engineering, Indian Institute of Technology, Kanpur, India

Abstract

This work proposes an efficient interactive melody adaptation method that improves melody extraction across domains using minimal annotated data. It selects regions for human annotation based on a confidence criterion and adapts the model using meta-learning to handle class imbalance.

Contributions

1. Novel meta-learning-based adaptation for class imbalance in classification.
2. Novel interactive domain adaptation method that combines active-learning with meta-learning.
3. New HAR dataset[a]

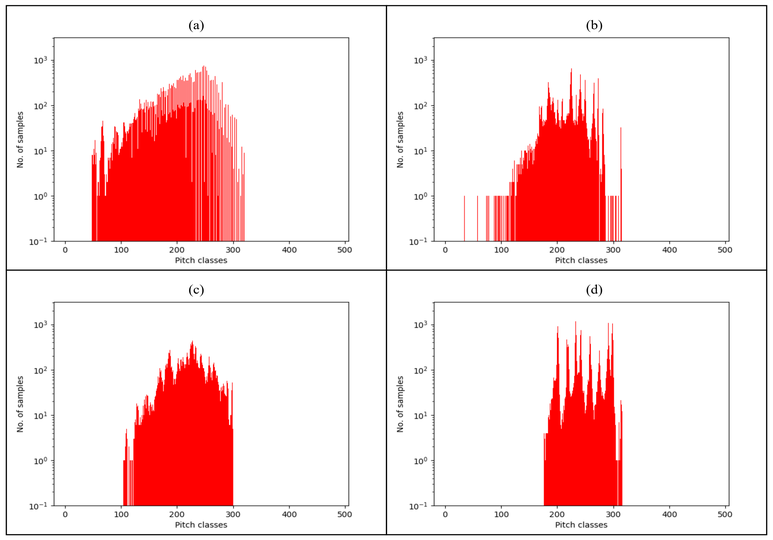

Figure 1. Class imbalance in (a) MIR1K (source), (b) ADC2004 (target), (c) MIREX05 (target), and (d) HAR(target).

Proposed Methodology

The magnitude spectrogram for the 5-second audio segments is computed using a short-time Fourier transform, employing a 2048-point (∼128ms) Hann window with a hop size of 10ms. Given the spectrogram as input, the model produces an output of shape $C \times M$, where each time frame m is categorized into one of $C = 450$ pitch classes. The pitch classes include a non-voiced pitch class and voiced pitch classes ranging from G#1 (51.91 Hz) to E6 (1318.51 Hz) with 1/8 semitone.

[a] https://zenodo.org/record/8252222

1. Step 1: Pretraining
   - $D_1^S = \{(X_i, Y_i)\}_{i=1}^{I}$; $X_i \in \mathbb{R}^{F \times T}$ and $Y_i \in \{0,1\}^{C \times M}$
   - Base model $f_{[\phi,\theta]}$ is trained on D_1^S as:

$$[f_{[\phi,\theta]}] \leftarrow [f_{[\phi,\theta]}] - \alpha \nabla_{[f_{[\phi,\theta]}]} L_w(f_{[f_{[\phi,\theta]}]}) \quad (1)$$

   The loss L_w is the weighted categorical cross-entropy loss given by:

$$L_w = -\sum_{i,c} w_c Y_{ic} \log(\hat{Y}_{ic}) \quad (2)$$

   Here, $w_c = \frac{1}{T_c}$; $T_c = \sum_{i,m,c} Y_{imc}$. Frozen ϕ, trainable θ.
2. Step 2: Confidence Model Training

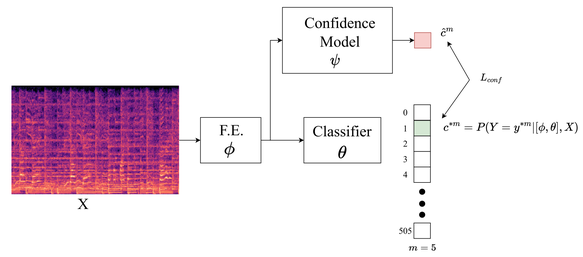

Figure 2. Here, ψ represents parameters of the confidence model. L_{conf} is calculated at a particular time frame $m = 5$.

3. Step 3: Active-Meta-Learning

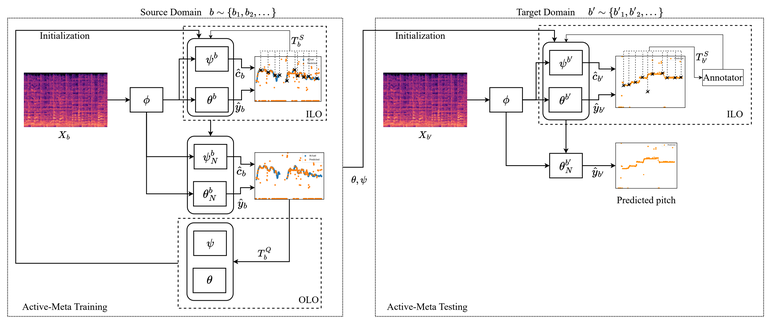

Figure 3. Framework of AML

- Consider another source training dataset $D_2^S = \{(X_b, Y_b)\}_{b=1}^{B}$
- Losses in ILO and OLO: weighted categorical cross-entropy loss, with weights as:

$$w_c'^g = w_c^g \times e^{\lambda |\Delta w_c|}, c = 0, 1, 2, \ldots, 505 \quad (3)$$

This is called as Meta Weighting (MW)

Results

Table 1. Performance metrics with the base model used by us and other baseline methods on the validation dataset (source dataset) and the three target datasets. All models are trained using CT. No adaptation is used.

Experiments	MIR1K-val		ADC2004		MIREX05		HAR	
	RPA	OA	RPA	OA	RPA	OA	RPA	OA
PB CNN	86.12	86.88	76.30	78.40	74.30	82.20	62.20	61.89
NMF-CRNN	88.48	88.97	74.45	74.73	75.78	76.10	63.19	63.40
Att. Net.	88.67	89.30	76.30	77.40	77.80	84.40	65.40	66.55
SegNet	**89.10**	**90.10**	**82.70**	**81.60**	78.40	78.60	68.34	65.63
Our base model	88.64	88.45	79.26	79.90	**81.88**	**81.30**	**75.43**	**75.90**

Table 2. Performance metrics with adaptive methods on the three target datasets. Here, MW, AA, and RA stand for meta-weighting, active adaptation, and random adaptation, respectively.

Experiments				ADC2004		MIREX05		HAR	
Method	MW	AA	RA	RPA	OA	RPA	OA	RPA	OA
FT-RA	-	✗	✓	80.34	80.98	81.16	82.10	76.45	76.88
FT-AA	-	✓	✗	81.55	81.66	81.80	81.85	76.95	76.17
MAML-RA	✗	✗	✓	81.10	81.41	83.16	83.28	77.70	78.10
MAML-AA	✗	✓	✗	81.32	81.99	82.12	81.80	75.78	75.98
w-AML(Ours)	✓	✓	✗	**86.40**	**86.15**	**87.23**	**87.80**	**80.60**	**81.45**

Table 3. Raw pitch accuracy on three target datasets for different support set size K.

Experiment	ADC2004			MIREX05			HAR		
	$K=10$	$K=15$	$K=20$	$K=10$	$K=15$	$K=20$	$K=10$	$K=15$	$K=20$
w-AML(Ours)	86.40	87.20	**88.95**	87.23	88.45	**89.99**	80.60	81.55	**82.01**

Table 4. Performance of the baseline method adapted by w-AML on the three target datasets.

Experiments	ADC2004			MIREX05			HAR		
	RPA	RCA	OA	RPA	RCA	OA	RPA	RCA	OA
NMF-CRNN	81.92	83.01	82.54	81.34	81.90	82.33	70.34	71.50	70.84

Google Scholar TASLP Paper

kavyars@iitk.ac.in