Audio-based Multimedia Event Detection

Name: Chethan Singh Mysore Jagadeesh

AndrewID: cmysorej

Problem Description:

Perform multimedia event detection(MED) with audio features.

Feature extraction:

Below is the list of feature extraction processes used:

1. MFCC feature extraction and 50 cluster Kmeans to get BoWs
2. SoundNet layer 10 is used to extract features and 50 cluster Kmeans to get BoWs
3. SoundNet layer 16 is used to extract features and 50 cluster Kmeans to get BoWs
4. Soundnet layer 16 is used to extract features and 40 cluster GMM to get BoWs
5. ASR features using TFIDF vectorization
6. Combine MFCC features and ASR TFIDF features
7. Combine features from SoundNet 16 layer-Kmeans BoWs with ASR TFIDF features
8. Combine features from SoundNet 16 layer-GMM BoWs with ASR TFIDF features

Data Preparation:

Since negative samples are much greater than positive samples, training set is prepared with all positive data points (~36) + (35 with NULL cases) + (18 with negative case 1) + (18 with negative case 2)

Training Process and validation results:

SVM was trained on each of the 8 features mentioned above with all of the below hyperparameter combinations for all 3 events:

kernel_type = ['linear', 'poly', 'rbf', 'sigmoid']

regparam = [0.01, 0.03, 0.1, 0.5, 1.0, 5.0, 10.0, 20.0, 40.0, 60.0, 80.0, 100.0, 110.0]

gamma_type = ['scale', 'auto']

Results with best SVM hyperparameters and along with AP for each event:

Sl. No	Feature type	P001	P002	P003
1	MFCC with 50 Kmeans BoWs	0.2166	0.3478	0.1219
		Kernel: rbf Regularization Param(C): 1.0 Gamma: scale	Kernel: rbf Regularization Param(C): 1.0 Gamma: scale	Kernel: rbf Regularization Param(C): 10.0 Gamma: scale
2	ASR with TFIDF vectors	0.1135	0.0922	0.3250
		Kernel: rbf Regularization Param(C): 20.0 Gamma: scale	Kernel: rbf Regularization Param(C): 0.03 Gamma: scale	Kernel: sigmoid Regularization Param(C): 5.0 Gamma: auto
3	MFCC + ASR	0.1565	0.1259	0.1741
		Kernel: sigmoid Regularization Param(C): 0.03 Gamma: scale	Kernel: sigmoid Regularization Param(C): 0.1 Gamma: scale	Kernel: linear Regularization Param(C): 0.5 Gamma: scale
4	Soundnet 10 with 50 Kmeans BoWs	0.2128	0.2757	0.1133
		Kernel: linear Regularization Param(C): 20.0 Gamma: auto	Kernel: rbf Regularization Param(C): 10.0 Gamma: scale	Kernel: poly Regularization Param(C): 5.0 Gamma: scale
5	Soundnet 16 with 50 Kmeans BoWs	0.1855	0.5157	0.1972
		Kernel: rbf Regularization Param(C): 0.5 Gamma: scale	Kernel: rbf Regularization Param(C): 0.5 Gamma: scale	Kernel: rbf Regularization Param(C): 10.0 Gamma: scale
6	Soundnet 16 with 50 Kmeans BoWs + ASR	0.2003	0.3879	0.3345
		Kernel: rbf Regularization Param(C): 10.0 Gamma: scale	Kernel: rbf Regularization Param(C): 100.0 Gamma: scale	Kernel: linear Regularization Param(C): 0.5 Gamma: scale
7	Soundnet 16 with 40 GMM BoWs	0.2565	0.4711	0.1824
		Kernel: rbf Regularization Param(C): 0.03 Gamma: scale	Kernel: rbf Regularization Param(C): 1.0 Gamma: scale	Kernel: rbf Regularization Param(C): 0.1 Gamma: auto
8	Soundnet 16 with 40 GMM BoWs + ASR	0.3011	0.2602	0.3768
		Kernel: rbf Regularization Param(C): 10.0 Gamma: scale	Kernel: rbf Regularization Param(C): 40.0 Gamma: scale	Kernel: rbf Regularization Param(C): 40.0 Gamma: scale

Best Scores and Models:

Sl.No	Event	Feature used	SVM hyperparams	AP	Overall AP
1	P001	Soundnet 16 with 40 GMM BoWs + ASR	Kernel: rbf Regularization Param(C): 10.0 Gamma: scale	**0.3011**	**0.3979**
2	P002	Soundnet 16 with 50 Kmeans BoWs	Kernel: rbf Regularization Param(C): 0.5 Gamma: scale	**0.5157**	
3	P002	Soundnet 16 with 40 GMM BoWs + ASR	Kernel: rbf Regularization Param(C): 40.0 Gamma: scale	**0.3768**	

Github:

https://github.com/ChetanMJ/LargeScaleMultiMedia