

Video-based Multimedia Event Detection

Name: Chethan Singh Mysore Jagadeesh

AndrewID: cmysorej

Problem Description:

Perform multimedia event detection(MED) with video features.

Feature extraction:

Below is the list of feature extraction processes used:

1. SURF features from OpenCV library with Hessian Threshold of 100. [Time taken 8+hrs].
After this 50 cluster Kmeans was used to create BoWs.[Time taken 1.5hrs]
2. All 1000 MobileNetV2 CNN features from last layer. [Time taken – 30mins]
3. 34 manually picked features from MobileNetV2 CNN last layer. [Time taken – 30mins]

Feature Index	Description
672	Mountain Tent
472	Cannoe
915	Yurt
656	Minivan
436	Beach wagon
518	Crash Helmet
580	GreenHouse
415	BakeHouse
449	BoatHouse
521	Crock Pot
572	Goblet
899	WaterJug
647	Measuring Cup
618	Ladle
469	Caldron
459	Mixing Bowl
438	Beaker
651	Microwave Oven
544	Dutch Oven
827	Stove
896	Washbasin
923	Plate

720	Pill Bottle
868	Tray
712	Petridish
898	Water Bottle
968	Cup
733	Pole
981	Baseball Player
693	Paddle
615	Kneepad
429	baseball

Data Preparation:

1. 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)
2. Further for Kfold validation was done on SVM model by including both Test and Train data. Even here data was balanced including all positive data points + (100 with NULL cases) + (25 with negative case 1) + (25 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	SURF with 50 Kmeans BoWs	0.1492	0.26492	0.1364
		Kernel: rbf Regularization Param(C): 1.0 Gamma: scale	Kernel: rbf Regularization Param(C): 5.0 Gamma: scale	Kernel: sigmoid Regularization Param(C): 0.03 Gamma: scale

2	MobileNetV2 34 features from last layer	0.7671	0.6527	0.5242
		Kernel: linear Regularization Param(C): 0.5 Gamma: scale	Kernel: rbf Regularization Param(C): 0.5 Gamma: scale	Kernel: rbf Regularization Param(C): 5.0 Gamma: auto
3	MobileNetV2 1000 features from last layer	0.6794	0.7451	0.4569
		Kernel: sigmoid Regularization Param(C): 0.1 Gamma: scale	Kernel: rbf Regularization Param(C): 5.0 Gamma: scale	Kernel: sigmoid Regularization Param(C): 0.3 Gamma: scale
4	MobileNetV2 34 features from last layer - Kfold	0.766	0.6262	0.4555
		Kernel: rbf Regularization Param(C): 0.01 Gamma: auto	Kernel: linear Regularization Param(C): 5.0 Gamma: scale	Kernel: rbf Regularization Param(C): 40 Gamma: scale
5	MobileNetV2 1000 features from last layer - Kfold	0.785	0.6339	0.5135
		Kernel: rbf Regularization Param(C): 20 Gamma: scale	Kernel: rbf Regularization Param(C): 40 Gamma: auto	Kernel: sigmoid Regularization Param(C): 5.0 Gamma: auto

Best Scores and Models:

Sl.No	Event	Feature used	SVM hyperparams	AP	Overall AP
1	P001	MobileNetV2 1000 features from last layer - Kfold	Kernel: rbf Regularization Param(C): 20.0 Gamma: scale	0.785	0.6812
2	P002	MobileNetV2 1000 features from last layer	Kernel: rbf Regularization Param(C): 5.0 Gamma: scale	0.7451	
3	P002	MobileNetV2 1000 features from last layer - Kfold	Kernel: sigmoid Regularization Param(C): 5.0 Gamma: scale	0.5135	

Github:

<https://github.com/ChetanMJ/LargeScaleMultiMedia>