Capturing the Temporal Domain in Echonest Features for Improved Classification Effectiveness

Alexander Schindler and Andreas Rauber

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Institute of Software Technology and Interactive Systems
Information and Software Engineering Group
Vienna University of Technology

Motivation

Advantages of the Million Song Dataset (MSD)

- Test algorithms on a large-scale collection
- Real-world scenarios
- Freely available
- Inter-linked to other data resources

Open Questions concerning the EN Features

- How do the EN features perform compared to other conventional feature sets?
- How to effectively aggregate the beat aligned vector sequences into a fixed length single vector representation?

The Echonest Features

as also provided by the MSD

Beat aligned Feature-Sequences

- Segments Timbre
- Segments Pitches
- Segments Loudness Max
- Segments Start

High-Level Features

- Key, Mode
- Tempo, Time Signature
- Energy, Danceability
- Song Hotttnesss

Unfortunately no reliable description of the extraction algorithms is provided.

Evaluation

Comparing the Echonest features against conventional feature sets

- Different combinations of EN features
- Compared by classification accuracy

Finding appropriate aggregation methods to convert vector sequences into fixed length single vectors

- Different combinations of statistical measures calculated from the beat aligned vector sequences

Conventional Featuresets

Set Name	Echonest Features	Dim	
Marsyas			Features from the Marsyas Framework developed by George Tzanetakis et al.
SPFE	Spectral Features	12	
MFCC	Mel-Frequency Cepstral Coefficients	52	
Chroma	Chroma Features	56	
Timb	Timbral Features	124	
Rhythm Patterns			A set of features by Rauber, Lidy et al. based on psychoacoustic models, capturing fluctuations on frequency bands critical to the human auditory system
RH	Rhythm Histograms	60	
SSD	Statistical Spectrum Descriptors	168	
TRH	Temporal Rhythm Histograms	420	
TSSD	Temporal Statistical Spectrum Descriptors	1176	
RP	Rhythm Patterns	1440	

Different Combinations of Echonest Features

Set Name	Echonest Features	Aggregators	Dim
EN0	Segments Timbre	mean	12
EN1	Segments Timbre	mean, variance	24
EN2	Segments Pitches	mean, variance	24
EN3	Segments Timbre	mean, covariance	90
EN4	Segments Timbre	mean, med, var, min, max, range, skewness, kurtosis	96
EN5	Segments Timbre, Segments Pitches	mean, median, var, min, max, range, skewness, kurtosis	192
Temporal Echonest Featues (TEN)	Segments Pitches, Segments Timbre, Segments Loudness Max, Segments Loudness Max Time, lengths of segments	mean, median, variance, min, max, range, skewness, kurtosis	216

Results

	Marsyas				Rhythm Patterns					Echonest Features						
Classifiers	**chrom**	**spfe**	**timb**	**mfcc**	**rp**	**rh**	**trh**	**ssd**	**tssd**	**EN0**	**EN1**	**EN2**	**EN3**	**EN4**	**EN5**	**TEN**
ISMIR Genre																
SVM Poly	50.3	54.9	67.7	62.1	75.1	64.0	66.5	78.8	80.9	67.0	75.1	64.3	67.2	78.5	80.4	81.1
KNN K1 L2	46.0	56.3	65.8	64.2	72.9	60.7	63.3	77.8	76.6	76.8	77.0	62.1	64.0	75.5	75.9	77.8
Rand-Forest	51.5	60.4	62.3	60.8	69.8	65.2	65.4	75.7	74.6	74.3	75.8	62.1	65.9	74.7	73.2	74.4
NaiveBayes	46.8	53.2	52.3	49.6	63.5	56.7	60.2	61.0	40.2	66.1	63.2	59.7	45.5	63.8	56.0	63.3
Latin Music Database																
SVM Poly	39.4	38.2	68.6	60.4	86.3	59.9	62.8	86.2	87.3	70.5	78.4	54.1	69.6	82.9	87.1	89.0
KNN K1 L2	37.3	42.5	62.7	58.4	74.3	58.7	49.5	83.1	78.4	73.5	78.7	57.1	52.2	77.3	79.0	80.9
Rand-Forest	39.4	46.4	58.1	53.6	58.8	50.3	47.5	76.3	73.0	69.9	74.7	53.3	54.9	74.1	73.5	75.9
NaiveBayes	26.9	35.7	43.5	46.7	66.0	47.0	49.9	64.1	67.8	66.5	68.4	47.0	40.4	70.8	71.1	73.3
GTZAN																
SVM Poly	41.1	43.1	75.2	67.8	64.9	45.5	38.9	73.2	66.2	56.4	61.1	37.0	53.6	63.9	65.2	66.9
KNN K1 L2	41.9	42.1	67.8	61.8	51.5	40.2	32.7	63.7	53.4	56.3	58.1	38.0	39.9	56.8	56.1	58.2
Rand-Forest	48.0	47.2	64.2	57.9	45.9	39.6	38.0	63.4	59.3	54.7	54.7	37.0	41.1	54.0	53.2	55.0
NaiveBayes	28.1	40.0	52.2	54.9	46.3	36.2	35.6	52.4	53.0	53.1	50.5	34.1	29.5	53.6	52.5	53.3
ISMIR Rhythm																
SVM Poly	38.1	41.4	60.7	54.5	88.0	82.6	73.7	58.6	56.0	55.1	63.1	38.7	51.7	62.7	63.7	67.3
KNN K1 L2	28.3	34.8	43.9	37.3	73.7	77.7	51.5	45.5	39.8	43.5	49.2	31.8	34.6	44.5	43.0	45.7
Rand-Forest	31.0	38.1	44.4	43.8	64.9	71.6	68.2	46.6	44.1	47.5	50.8	35.2	37.1	47.9	48.8	53.5
NaiveBayes	23.3	37.0	37.7	36.5	75.9	69.0	69.3	44.4	46.8	52.8	53.3	39.7	25.1	52.8	49.9	55.1

Results show that

- Echonest features perform well compared to conventional feature sets
- Already simple combinations of features and statistical measures lead to acceptable results
- Harnessing the temporal domain of the beat aligned vector sequences provides results outperforming conventional feature sets on “traditional” benchmark sets

Recommendations

EN0 - EN1
- Provide acceptable results
- Short feature vectors
- Recommended for applications focusing on runtime behavioural aspects

EN4-EN5
- Provide good results
- Higher number of dimensions
- An acceptable compromise between low dimensional EN0-EN1 and high dimensional TEN

TEN
- Provide often results outperforming conventional feature sets
- Recommended for applications focusing on accuracy

Resources

We provide a number of benchmark partitions that researchers can use in their future studies, in order to facilitate repeatability of experiments with the MSD beyond x-fold cross validation. We also encourage and provide a platform for exchange of results obtained and new partitions created via our web site:

`http://www.ifs.tuwien.ac.at/mir/msd/`

- All Echonest feature combinations presented by this evaluation including Temporal Echonest Features
- Additional feature sets for the Million Song Dataset extracted from 99.5% of downloaded audio samples
- Ground truth assignments for 42% of the Million Song Dataset downloaded from Allmusic.com
- Splits with all the ground truth assignments into genre and style classes, artist or album filters, with “traditional” partitioning into train/test splits as well as stratified splits.

10th international workshop on Adaptive Multimedia Retrieval
24-25 October 2012 - Copenhagen, Denmark