ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 4, 2024

https://zenodo.org/records/14506452

Developing Audio Features for Real-Time Genre Recognition in Music

Ramesh Gugulothu[1], Assistant Professor[1], Department of ECE, Siddhartha Institute of Technology & Sciences, Telangana, India

Nalla Vijay Kumar[2], Assistant Professor[2], Department of CSE, Siddhartha Institute of Technology & Sciences, Telangana, India.

Abstract

The scenarios opened by the increasing availability, sharing and dissemination ofmusic across the Web is pushingfor fast, effective and abstract ways oforganizingand retrieving music material. Automatic classification is a central activity to modelmost of these processes, thus its design plays a relevant role in advanced MusicInformation Retrieval. In this paper, we adopted a state-of-the-art machine learning algorithm, i.e. Support Vector Machines, to design an automatic classifier of musicgenres. In order to optimize classification accuracy, we implemented some alreadyproposed features and engineered new ones to capture aspects of songs that havebeen neglected in previous studies. The classification results on two datasets suggestthat our model based on very simple features reaches the state-of-art accuracy (onthe ISMIR dataset) and very high performance on a music corpus collected locally.

Introduction

Music genres are difficult to describe as there is no complete agreement on their definition."*Genres emerge as terms and nouns that define recurrences and similarities that membersof a community make pertinent to identify musical events*" (Fabri, 1997)The notion of community corresponds to a complex self-organizing system that triggersthe development and assessment of a *genre*. In this perspective, the community plays therole of an ontology designer which implicitly defines properties and rules of the targetgenre as well as its differences with external habits and trends.Given the high complexity of such system, to define a model for automatic genreclassification, we should capitalize from the work carried out in Information Retrieval (IR).This has shown that document relevance with respect to a user's query (e.g. a particularsong) is not determined by only local properties, e.g. the query and the retrieved items,as global notions, that emerge from the entire corpus, are also important. Indeed, everyquantitative model in IR relies on a large number of parameters (e.g.*term weights*) that depend on the set of *all* indexed documents. In order to model a musical genre, it isthus critical to study local (the target genre examples) and global (the examples of othergenres) characteristics and express them in term of statistical properties.

Such concepts are the foundations of modern machine learning algorithms (Mitchell,1997) which aim to model classification functions based on the sets of positive and negativeexamples, i.e. the songs that belong or not to a target genre. As the machine learningapproaches are quite standard and they tend to behave similarly on different applicationdomains, the actual complexity relates mainly to the feature design task. The role offeatures is to provide a description of example songs that can be processed by learningalgorithms. These will guide the induction of the classification function in agreement withsuch descriptions. As we would like to classify songs stored as audio files, i.e. waveforms,the design of features is quite complex and requires the application of signalanalysistechniques.

In this paper, we experimented a state-of-the-art machine learning algorithm, i.e.Support Vector Machines, in the design of an automatic genre classifier over audio infor-mation. In order to optimize the classification accuracy of our model, we implementedsome features described in literature and designed new features to capture aspects pre-viously neglected. We experimented our models on annotated collections (i.e. classifieddata instances) made available in previous investigation (*Magnatune* dataset) as well ason a novel data collection, designed to carry out a cross-collection comparison. The results obtained on large scale experiments suggest that our model based on very simple featuresreaches the state-of-art accuracy on the *Magnatune* dataset and very high performanceon our new music corpus.

The remainder of this paper is organized as follows: Section 2 introduces SupportVector Machine and kernel methods, Section 3 describes the basic and new set of features,Section 4 shows the experiments

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 4, 2024

https://zenodo.org/records/14506452

with music genre categorization and finally Section 5summarizes the conclusions.

Related Work

The description of the basic features used in ourexperiments can be found in (Tzanetakiset al., 2001). The main idea that we inherited from Tzanetakis et al. is the split between

superficial features, called *musical surface* and advanced features, i.e.*rhythm features*.The literature experiments show that such baseline features achieve very high accuracy.Another very inspiring work is (Tzanetakis et al., 2002) in which the concept of beat strength is defined as a rhythmic characteristic that allows us to discriminate betweentwo pieces having the same tempo.

The *MFCC* feature has been used in music categorization in (Logan, 2000). Such studydemonstrates the importance of using *MFCC* for music classification. Moreover, MFCCwas used in the Pampalk's system that won the MIREX 2004competition. In such work,the feature extraction method was based on a frame cluster similarity (Pampalk, 2005).*MFCC* was also used in another system based on AdaBoost (Bergstra and Casagrande,

2005) that won MIREX 2005.In (Lidy and Rauber, 2005), a combination of features based on rhythm patterns,statistical descriptors and rhythm histograms was used. In (Gouyon et al., 2004), it wasconsidered a specific set of rhythmic descriptors for which was provided procedures ofautomatic extraction from audio signals. The corpus used in our experimentation wasalso used in (Pampalk et al., 2005), with a set of *spectral features*, e.g.*MFCC*, and a setof advanced features, called *fluctuation patterns*.In (Berenzweig et al., 2003), it is described a method of music mapping into a semanticspace that can be used for music similarity measurement. The value along each dimensionof this *anchor space* is computed as the output from a pattern classifier which is trainedto measure a particular semantic feature. In anchor space, distributions that representobjects such as artists or songs are modeled with Gaussian Mixture Models. An interestingapproach is used in (Mandel et al., 2005) where it is described a system for performingflexible music similarity queries using SVM active learning. In (Lippens et al., 2004) itis shown that, although there is room for improvement, genre classification is inherentlysubjective and therefore perfect results can not be expected neither from automatic norfrom human classification.

Genre Classifier based on Support Vector Machines

Many learning algorithms consider features as dimensions of a vector space. Each in-stance is represented by a feature vector where the components are the numeric valuesassociated with features. Support Vector Machines (SVMs) (Vapnik, 1995) are state-of-

the-art learning methods based on vector spaces. One of their interesting properties isthe possibilities of using kernel functions. These allow SVMs to implicitly generate largefeature spaces like for example the space of feature conjunctions. The next section brieflyintroduces this interesting machine learning approach.

Support Vector Machines

To apply SVMs to music classification, we need a function *Á* :*S !* <*n* to map our songspace *S* into <*n*. Given such vector space and a set of positive and negative examplesmapped in vectors, SVMs classify them according to a separating hyperplane, *H*(~*x*) = ~*w ¢* ~*x* + *b* = 0, where ~*x* = *Á*(*s*)*; s 2 S* and the two parameters ~*w 2* <*n* and *b 2* <(learnedby applying the *Structural Risk Minimization principle* (Vapnik, 1995)). More in detail,they are learned by solving the following optimization problem:

$$\begin{cases} min \quad ||\vec{w}||^2 + C\sum_{i=1}^{m}\xi_i^2 \\ y_i(\vec{w}\cdot\vec{x_i}+b) \geq 1-\xi_i, \quad \forall i = 1,..,m \\ \xi_i \geq 0, \quad i = 1,..,m \end{cases} \tag{1}$$

where ~*xi* are the training instances, *m* is the number of such instances and *»i*are the slackvariables of the optimization problem. From the kernel theory we havethat:

$$H(\vec{x}) = \left(\sum_{i=1..l} y_i\alpha_i\vec{x}_i\right)\cdot\vec{x}+b = \sum_{i=1..l} y_i\alpha_i\vec{x}_i\cdot\vec{x}+b = \sum_{i=1..l} y_i\alpha_i\phi(s_i)\cdot\phi(s)+b = 0,$$

where *yi* is equal to 1 for a positive example and or -1 for a negative example, ®*i 2* <with®*i* ¸ 0, *Á*(*si*) = ~*xi* *8i 2 f*1*; ::; lg* are the training instances and the

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 18, Issue 4, 2024
https://zenodo.org/records/14506452

product $K(si; s) = h\acute{A}(si) ¢ \acute{A}(s)i$is the kernel function associated with the mapping $\acute{A}$. The simplest mapping
that we can apply is $\acute{A}(s) = \sim x = hx1; :::; xni$where $xi= 1$ *if* the feature *i*appears in thesong *s* otherwise $xi = 0$. If we use as a kernel function the scalar product, we obtain thelinear kernel $KL(si; s) = \sim xi ¢ \sim x$.

Another interesting kernel is the polynomial one, i.e.$Kp(si; s) = (c + \sim xi ¢ \sim x)d$, wherec is a constant and *d* is the degree of the polynom (Basili and Moschitti, 2005). Thepolynomial kernel is equivalent to carry out the scalar product in the space of featureconjunctions, where the number of features in each conjunction is up to *d*. For example,if we have two features such as *pitch* and *volume*, the learning algorithm can test if somecombinations characterizes a particular genre, e.g. the combination of low *pitch* andlow *volume* is typical of *classical music*.Although, kernel methods are a powerful toolsto learnclass differences, it is very important to define a *good*set of features achievingoptimal results.

Extracting Features from Audio Files

In this paper, we use several basic features proposedin music classification literature andwe also propose new interesting ones. The next sections are devoted to the description ofthe experimented features.

Simple or Basic Features

We represent the musical surface of each song by means of the statistics of the spectraldistribution over time. In particular, we analyze the average and standard deviation of 6-dimensional vectors over the entire song. Such dimensions, *volume*, *beats*, *spectral energy*,*centroid*, *pitch* and *5-MFCC*, are described hereafter.

- *Volume*. Given N song samples, S_k, the *Volume* is:

$$Volume = \sqrt{\frac{1}{N}\sum_{k=1}^{N} A(S_k)^2} \qquad (2)$$

where Sk is one of the samples stored in the buffer and $A(Sk)$ is the amplitude ofthe signal at time Sk. This function is not equivalent to the concept of volumeused in signal processing, but it gives higher values for louder sounds and viceversa,which is enough for our purposes. Moreover, during the preprocessing phase, wenormalize the volume of each song as we are not interested to absolute values butto the relative differences between two frames.

The *Spectral Energy* is correlated to the Fourier Transform, which maps audio sig-
nal into frequency domain. For each audio sample set, we compute Fast FourierTransformation (FFT). *Centroid* is an interesting psycho-acoustical feature that measures the mean spectralfrequency in relation with the amplitude; in other words the position in Hz of thecenter of mass of the spectrum. it is useful as a measure of the sound brightness.

$$C = \frac{\sum_{k=0}^{N/2} f_k |X(k)|}{\sum_{k=0}^{N/2} |X(k)|} \qquad (3)$$

where N is the FFT size, $X(k)$, $k = 0, .., N$ is the FFT of the input signal, and f_k, $k = 1, .., N$, is the k-th frequency bin.

Pitch is the perceived fundamental frequency of a sound. This can be computedby an autocorrelation algorithm applied to audio signals. We defned a different
approximation of the above notion as it captures more information. We definechroma vectors as 12-element vectors, where each component represents the spectral
energy corresponding to one pitch class (i.e. C, C#, D, D#, etc.). The algorithm,builds the 12 chroma vectors by deriving components from the main frequencies
in a temporal frame. Durations below some threshold's are not taken into accountas they are considered not meaningful for the frame. Notes corresponding to eachfrequency are then mapped according to a fuzzy matching based on a reference
octave frequency: in order to discretize, i.e. select the proper note, the frequenciesof an *A* note in every octave (for a total of 8 octaves) are taken as reference.

The *Mel Frequency Cepstral Coefficients* (MFCCs) are well known compact formsthat can represent speeches. They are the most common representation used for*Spectra* in Music Information Retrieval (MIR). The following is a brief algorithmfor their computation:

1. apply window function;
2. compute power spectrum (using FFT);
3. apply Mel filter bank;
4. apply Discrete Cosine Transform (DCT);

The MFCCs have important advantages: they are simple and fast, well tested.Moreover, they have also

ISSN 2454-9940
www.ijasem.org
Vol 18, Issue 4, 2024
https://zenodo.org/records/14506452

a compressed and flexible (i.e. easy to handle) representation (Logan, 2000).

Complex, Synchronous and Structural Features

In this section we describe our new designed features.

Beats

The *Beats* feature tries to count the number of beats of a song. Generally, beats aredriven by instruments that operate in the lower frequencies, like the drum or the bass.In order to obtain a feature able to count the number of beats, we apply a lowpass tothe target song. This will cut o® frequencies higher than 200 Hz ((McNab et al., 1996),(Marolt, 2006), (Davies and Plumbley, 2005)). Most instruments playing in a song will beattenuated or totally eliminated. The remaining sounds are usually related to the drumsand the bass.

More in detail, our algorithm uses the volume feature value together with 10 low-passfilters that cut over the frequencies higher than 200 Hz. We use a bank of ten filtersin order to minimize the sound distortion and the computation time. Then, the songwavelength is discretized to analyze the *attack* of each beat (referred as *peak duration*
below), i.e. the time that elapses between the start of apeak and the successive silentphase. As the temporal range of the target frame is fixed the number of peaks is alsoinformative about the rhythm (i.e. it is a crude but useful approximation of the notionof *bpm* local to a sample). An example of this process is given in Figure 1. Given suchdiscretized wave, we extract five values:

- average and variance of peak duration
- average and variance of peak distance
- average beats per minute

The figure 2 shows two music moments that canclearly be distinguished using the beatsfeature. In the first moment, a series of regular impulses caused by a drum is presentwhereas in the second we find a more complex texture produced by a bass. Moreover,analyzing the amplitude of a wave, we can determinethe classification of genres: songsof classical and jazz genres show lower waves contrarily to rock and electronic songs.

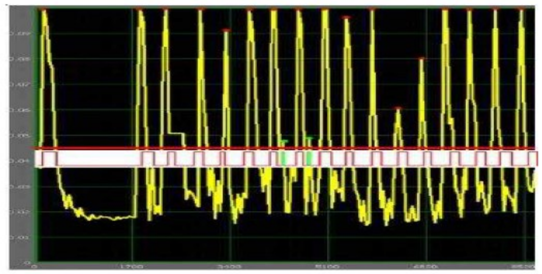

Figure 1: An example of beats for electronic music

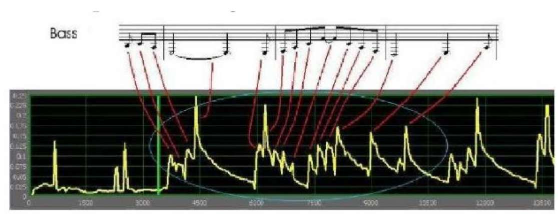

Figure 2: An example of beats for jazz-blues music

Volume Reverse

The intuition about this feature come from examining song recording methods. Thefirst step in a recording process is to collect the sounds of every instrument. Severalmicrophones can be used for the same instrument or, viceversa, the same microphone can
be used for several instruments. When the recording phase of a single track is completed, amaster multi-track is mixed in stereo channels so that a song can be played by conventionalhifiequipments. For example, rock, pop and electronic music, is often produced bymoving a sound of one instrument from one stereochannel to the other (sound effects likeecho and surround). Moreover, such music is enriched with sonorous effects. For example,with rock music guitar distortions are often used to make the sound less uniform whereaswith pop and electronic music scratch effect is applied.

The above techniques make the audio wave of a channel very different from the other.On the contrary, classic and jazz music is recorded with different modalities. First of all,instruments' distortions and sonorous effects are quite rare because this music is basedon a cleaner type of sound. The recording technique is direct and makes a large use ofenvironment microphones as it is preferred toemphasize live recording, giving much moreimportance to the solos and improvisations.

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 4, 2024

https://zenodo.org/records/14506452

Thisproduces a stereo track with two verysimilar channels.

By considering the different recording methods, we can distinguish rock, pop andelectronic music from classic and jazz. For this purpose, we designed a feature thatmeasures the variation between the sound wave of the two stereo channels. More in detail,we subtract the audio wave of a channel to the other oneand compute the absolute value.We expect that for classic and jazz music the values tend to be around zero whereas for

Features	*Multiclassifier*	*Rock*	*Classic*	*Jazz*	*Electronic*	*Metal*	*World*
	Accuracy	F1 measure					
Basic	80.5	0.63	0.93	0.68	0.76	0.59	0.73
Basic + Pitch	80.0	0.61	0.93	0.65	0.76	0.56	0.72
Basic + Beats	80.4	0.64	0.93	0.6	0.8	0.57	0.7
Basic + Chorus	80.6	0.62	0.93	0.68	0.77	0.58	0.73
Basic + Reverse	80.9	0.62	0.94	0.72	0.77	0.58	0.74
Basic + Reverse + Chorus	82.0	0.66	0.94	0.7	0.79	0.61	0.74
Basic + Reverse + Chorus + Pitch	81.9	0.64	0.94	0.7	0.80	0.58	0.74
Basic + Reverse + Chorus + Beats	81.2	0.62	0.93	0.68	0.82	0.58	0.71
Basic + All Advanced	82.3	0.66	0.94	0.68	0.83	0.61	0.72
Pampalk 2004	84,10	-	-	-	-	-	-

Table 1: Accuracy on the Magnatune 2004 Corpus

Features	*Multiclassifier*	*Rock*	*Classic*	*Jazz*	*Electronic*	*Pop*
	Accuracy	F1 measure				
Basic	89.4	0.85	0.93	0.9	0.91	0.87
Basic + Pitch	90.8	0.89	0.93	0.9	0.91	0.9
Basic + Beats	89.8	0.85	0.94	0.9	0.92	0.86
Basic + Chorus	88.4	0.83	0.93	0.9	0.91	0.85
Basic + Reverse	90.0	0.86	0.94	0.91	0.91	0.88
Basic + Reverse + Chorus	90.8	0.87	0.94	0.91	0.92	0.89
Basic + Reverse + Chorus + Pitch	91.6	0.9	0.94	0.9	0.92	0.92
Basic + Reverse + Chorus + Beats	91.4	0.88	0.95	0.93	0.91	0.89
Basic + Advanced (ALL)	92.0	0.89	0.96	0.92	0.91	0.91

Table 2: Accuracy on the RTV Corpus

electronic, rock and pop values are subject to sudden changes.

Chorus

A characteristic of rock and pop songs is the presenceof a periodic structure: to a versefollows a chorus and so on until the end of the song. It is also applied a change of tonalityat the end of the song. This schema is less strict in the electronic, jazz and classics songs.The latter two musical genres show moreimprovisation and the presence of very technicalsolos which make songs much more complex and less rigidin their internal structure.Finding a general schema of the songs can help to distinguish between jazz, classic andelectronic genre from music much closer to rock and pop.

Our algorithm to detect such schema eliminates thevoice of the singer (if there isany), to preserve only the audio data given by musical instruments. This is carried outby subtracting the wave of the two audio channels (of course, this can be done only if theanalyzed song is stereo). Such approach will eliminatethe middle channel audio whichgenerally is the part containing the singer voice.

After the above step, a schema of the song can bedetected based on the analysis ofthe spectral energy. In particular, we have noted that the chorus of pop or rock music isassociated with greater values of energy. Thus, our algorithm computes mean and meansquare values of the audio wave of the frequency of thedetected chorus part.

Experiments

In these experiments, we tested our SVM song classifier on two different data sets, *Magnatune 2004* and a novel music corpus (RTV) that contains some *Magnatune*songs mixedwith commercial music. We experimented with different feature combinations to studytheir usefulness in characterizing different genres.

Experimental Setting

The *Magnatune 2004* corpus is composed by 729 songs distributed in 6 genres as follows*Rock* 13.7%, *Classic* 44.2%, *Jazz* 3.6%, *Electronic* 15.8%, *Metal* 6.2% and *World* 16.6%.All the songs are free from copyrights and can be downloaded from http://ismir2004.ismir.net/. This corpus is quite difficult to classify for at least three reasons: (1) the *World* genre classification is quite complex as it encloses several different sounds andstyles; (2) there is a strong similarity between music *Rock* and *Metal*; and (3) there arevery few examples of *Jazz*.

The RTV corpus is composed by 500 songs selected from *Magnatune* and some songs selected by proprietary databases1. Such songs areequally distributed on 5 genres (eachof them contains 100 songs): *Rock*, *Classic*, *Jazz-blues*, *Electronic* and *Pop*. *Rock* and *Pop*classes are composed by commercial music (e.g. Madonna and Depeche Mode for Popand Metallica and Korn for Rock). The songs of the other classes are randomly selectedfrom *Magnatune*.

For the experiments, we used the WEKA software available at http://www.cs.waikato.ac.nz/»ml/weka/. We used the default parameters and the polynomial kernel (ofWaikato,2006). We trained an SVM for each class in the scheme ONE-vs-ALL (Rifkin and Klautau, 2004). For each testing instance, we selected the class associated with the highestSVM score. The classification performance of the individual class is

ISSN 2454-9940
www.ijasem.org
Vol 18, Issue 4, 2024
https://zenodo.org/records/14506452

evaluated with the$F1$ measure. This assigns equal importance to Precision P and Recall R, i.e.$f1 = 2P£R$ $P+R$ .The multiclassifier performance is measured by means of accuracy.

Classification Results

For both corpora we applied a 10-fold cross validation.This means that we divided eachcorpus in 10 parts and 9 of them were used for training and 1 for testing. By rotating thetesting sample, we obtained 10 different measures on which we evaluated the average.

Table 1 reports the results for the *Magnatune*collection. Column 1 shows the featuresets used to represent the song instances, Column 2 reports the multiclassifier accuracy,and the columns from 3 to 6 illustrates the $F1$ measures of the *Rock*, *Classic*, *Jazz*, *Electronic*, *Metal* and *World* binary classifiers,respectively. We note that the morefeatures we use the higher the multiclassifier accuracy is. Indeed, the best result is achievedusing the basic features plus the new ones, i.e. 82.3%. Consequently, the new featuresimprove the basic features of about 2 absolute percentpoints. This enhancement is notneglectable since it is difficult to improve an already high baseline, i.e.80.45%. Moreover,82.3% is very near to the best figure obtained on the *Magnatune* corpus, i.e. the 84.1%derived in (Pampalk, 2005). Note that the features used to obtain such state-of-the-artaccuracy are remarkably more complex to extract than those proposed in our model. Suchcomplexity made difficult to implement and study a model that combines such featureswith those that we propose. Although this will be part of our future work.

Regarding the individual categories, we observed from the confusion matrix that *Rock* is often misclassified in place of *Metal* and viceversa. The *Jazz* classifier has a low accuracy; this suggests that it is difficult to recognize *Jazz*songs. An alternative explanation is1As these songs are protected by copyrights, we could not make RTV available but we are going toprovide the learning files in WEKA format.

the low number of training instances available to train the corresponding binary classifier.On the contrary, the accuracy on *Classic* and *Electronic* genres is quite high. This can beexplained by the remarkable differences in terms of musicological and sonorous aspects.

With the aim of showing that our features capture important difference between thediverse genres, we experimented our classifiers on the RTV collection. The results arereported in Table 2, which is very similar to the previous Table except for the presence of the Pop category. We note that the accuracy is in general much higher than the oneobtained on the *Magnatune*test set. The main reason is that in RTV each category hasan enough number of positive examples for training (i.e. 100). This does not happen for*Magnatune*. For example, *Jazz* has only 26 training songs. Moreover, we still observe animprovement of about 2% of the classification accuracy when the new features are addedto the basic ones.

In particular we empathize the relevance of the feature *Volume Reverse* in the classification results related to the *jazz* and *world*. When this feature is added to the set thef-measure of this genre reaches the peak.
Finally, it should be noted that also the accuracy obtained with the basic featuresis very high on both collections. This is due to the use of (a) a powerful learning algorithm, i.e. SVMs, and (b) the polynomial kernel that generates many interesting feature conjunctions.

Conclusions

The large availability of songs across the Web requires efective ways of automaticallyorganizing and retrieving music material. Automated genre classification is thus a criticalstep to carry out such processes.

In this paper, we adopted a state-of-the-art machine learning algorithm, i.e. Support Vector Machines, to design an automatic classifier of music genres. To improve the classification accuracy of our system, we used previous designed features and we engineered new ones. The classification accuracy on two datasets show that our model based on very simple features approaches the state-of-art systems. This good result is due to both ournovel features and the use of a powerful learning model, i.e. SVMs along with the verypromising techniques based on kernel methods.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 4, 2024

https://zenodo.org/records/14506452

References

R. Basili and A. Moschitti. Automatic Text Categorization: from Information Retrieval to Support Vector Learning. Aracne Publisher, 2005.

A. Berenzweig, D. Ellis, and S. Lawrence. Anchor space for classi¯cation and similarity measurement of music. In Proceedings of IEEE ICME, 2003.

J. Bergstra and N. Casagrande. Two algorithms for timbre and rhythm-based multi-resolution audio classi¯cation. In Proceedings of ISMIR, 2005.

M. E. P. Davies and M. D. Plumbley. Beat tracking with a two state model. In Proceedings of IEEE ICASSP, 2005.

F. Fabri. Browsing music spaces: Categories and the musical mind. Website, 1997. URL http://www.mediamusicstudies.net/tagg/others/ffabbri9907.html.

Conference RIAO2007, Pittsburgh PA, U.S.A. May 30-June 1, 2007 - Copyright C.I.D. Paris, France

F. Gouyon, S. Dixon, E. Pampalk, and G. Widmer. Evaluating rhythmic descriptorsfor musical genre classification. In Proceedings of AES 25th International Conference, 2004.

T. Lidy and A. Rauber. Combined fluctuation features for music genre classification. InProceedings of MIREX, 2005.

S. Lippens, J.P. Martens, T. De Mulder, and G. Tzanetakis. A comparison of human andautomatic musical genre classification. In Proceedings of the International Conferenceon Acoustics, Speech and Signal Processing (ICASSP), 2004.

B. Logan. Mel frequency cepstral coefficients for music modeling. In Proceedings of ISMIR,2000.

M. Mandel, G. Poliner, and D. Ellis. Support vector machine active learning for musicretrieval. ACM Multimedia System Journal, 2005.
M. Marolt. A mid-level melody-based representation for calculating audio similarity. InProceedings of ISMIR, 2006.

R.J. McNab, L.A. Smith, I.H. Witten, C.L. Henderson, and S.J. Cunningham. Towardsthe digital music library: Tune retrieval from acoustic input. In Proceedings of the ACMConference on Digital Libraries, 1996.

T. M. Mitchell. Machine Learning. Mc Graw Hill, 1997.University of Waikato. Weka, a platform for automatic classification. http://www.cs.waikato.ac.nz/ ml/weka/.Website,2006.URLhttp://www.cs.waikato.ac.nz/ ml/weka/.

E. Pampalk. Speeding up music similarity. In Proceedings of MIREX, 2005.E. Pampalk, A. Flexer, and G. Widmer. Improvements of audio-based music similarityand genre classificaton. In Proceedings of ISMIR, 2005.

R. Rifkin and A. Klautau. In defense of one-vs-all classification. J. Mach. Learn. Res., 5: 101{141, 2004. ISSN 1533-7928.

G. Tzanetakis, G. Essl, and P. Cook. Automatic musical genre classification of audiosignals. In ProceedingsofISMIR,2001.URLgtzan@cs.princeton.edu.

G. Tzanetakis, G. Essl, and P. Cook. Human perception and computer extraction ofmusical beat strength. In Procedings of Digital Audio E®ects (DAF), 2002.

V. Vapnik. The Nature of Statistical Learning Theory. Springer, 1995.