Music Data Mining: An Introduction

Tao Li and Lei Li

October 17, 2010

Contents

Chapter 1

Introduction

In the Internet age, a gigantic amount of music-related information is easily accessible. For example, music and artist information can be obtained from the artist and record company web sites, song lyrics can be downloaded from the lyrics databases, and music reviews are available from various discussion forums, blogs, and on-line magazines.

As the amount of available music-related information increases, the challenges of organizing and analyzing such information become paramount. Recently, many data mining techniques have been used to perform various tasks (e.g., genre classification, emotion and mood detection, playlist generation, and music information retrieval) on music-related data sources. Data mining is a process of automatic extraction of novel, useful and understandable patterns from a large collection of data. With the large amount of available data from various sources, music has been a natural application area for data mining techniques. In this chapter, we attempt to provide a review of music data mining by surveying various data mining techniques used in music analysis. The reader should be cautioned that music data mining is such a large research area that truly comprehensive surveys are almost impossible, and thus, our overview may be a little eclectic. An interested reader is encouraged to consult with other papers for further reading, in particular, [Jensen, 2008, Orio, 2006]. In addition, one can visit the website http://users.cis.fiu.edu/~lli003/Music/music.html, in which a comprehensive survey on music data mining is provided and will be updated constantly.

1.1 Music Data Sources

Table 1.1 briefly summarizes various music-related data sources, describing different aspects of music. We also list some popular websites, from which music data sources can be obtained. These data sources provide abundant information related to music from different perspectives. In order to better understand the data characteristics for music data mining, we give a brief introduction to various music data sources below.

Data Sources	Examples (websites)
Music Metadata	*All Music Guide*, *FreeDB*, *WikiMusicGuide*
Acoustic Features	*Ballroom*
Lyrics	*Lyrics*, *Smartlyrics*, *AZlyrics*
Music Reviews	*Metacritic*, *Guypetersreviews*, *Rollingstone*
Social Tags	*Last.fm*
User Profiles and Playlists	*Musicmobs*, *Art of the Mix*, *Mixlister*
MIDI Files	*MIDIDB*, *IFNIMIDI*
Music Scores	*Music-scores*

Table 1.1: Various Music Data Sources.

Music Metadata: Music metadata contains various information describing specific music recordings. Generally speaking, many music file formats support a storing structure known as *ID3*, in which music metadata, such as artist name, track title, music description and album, will be stored. Another way for accessing music metadata is to use online music metadatabase applications. Metadatabase applications are used by the majority of music softwares for the purpose of providing descriptions for CD trackings. Some well-known music metadatabase applications, *e.g.*, *All Music Guide* and *FreeDB*, provide flexible platforms for music enthusiasts to search, upload and manage music metadata.

Acoustic Features: Music acoustic features include any acoustic properties of an audio sound that may be recorded and analyzed. For example, when a symphonic orchestra is playing Beethoven's *9th Symphony*, each musical instrument, with the exception of some percussions, produces different periodic vibrations. In other words, the sounds produced by musical instruments are the result of the combination of different frequencies. Some basic acoustic features [Orio, 2006] are listed in Table 1.2.

Lyrics: Lyrics are a set of words that make up a song in a textual format. In general, the meaning of the content underlying the lyrics might be explicit or implicit. Most lyrics have specific meanings, describing the artist's emotion, religious belief, or representing themes of times, beautiful natural scenery and so on. Some lyrics might contain a set of words, from which we cannot easily deduce any specific meanings. The analysis of the correlation between lyrics and other music information may help us understand the intuition of the artists. On the Internet, there are a couple of websites offering music lyrics searching services, *e.g.*, *SmartLyrics* and *AZLyrics*.

Music Reviews: Music reviews represent a rich resource for examining the ways that music fans describe their music preferences and possible impacts of those preferences. With the popularity of Internet, an ever-increasing number of music fans join the music society and describe their attitudes towards mu-

Acoustic Features	Description
Pitch	related to the perception of the fundamental frequency of a sound; range from low or deep to high or acute sounds.
Intensity	related to the amplitude, of the vibration; textual labels for intensity range from soft to loud.
Timbre	defined as the sound characteristics that allow listeners to perceive as different two sounds with same pitch and same intensity.
Tempo	the speed at which a musical work is played, or expected to be played, by performers. The tempo is usually measured in beats per minute.
Orchestration	due to the composers' and performers' choices in selecting which musical instruments are to be employed to play the different voices, chords, and percussive sounds of a musical work.
Acoustics	a specialization on some characteristics of timbre, including the contribution of room acoustics, background noise, audio post-processing, filtering, and equalization.
Rhythm	related to the periodic repetition, with possible small variants, of a temporal pattern of onsets alone. Different rhythms can be perceived at the same time in the case of polyrhythmic music.
Melody	a sequence of tones with a similar timbre that have a recognizable pitch within a small frequency range.
Harmony	the organization, along the time axis, of simultaneous sounds with a recognizable pitch.

Table 1.2: Different Acoustic Features.

sic pieces. Online reviews can be surprisingly detailed, covering not only the reviewers' personal opinions but also important background and contextual information about the music and musicians under discussion [Hu et al., 2005].

Music Social Tags: Music social tags are a collection of textual information that annotate different music items, such as albums, songs, artists and so on. Social tags are created by public tagging of music fans. Captured in these tags is a great deal of information including music genre, emotion, instrumentation, and quality, or simple description for the purpose of retrieval. Music social tags are typically used to facilitate searching for songs, exploring for new songs, finding similar music recordings, and finding other listeners with similar interests [Lamere, 2008]. An illustrative example of well-known online music social tagging systems is *last.fm*, which provides plenty of music tags through public tagging activities.

User Profiles and Playlists: User profile represents the user's preference to music information, *e.g.*, what kind of songs one is interested in, which artist one likes. Playlist, or called listening history, refers to the list of music pieces that one prefers or has listened to. Traditionally, user profiles and playlists are stored in offline music applications, which can only be accessed by a single user. With the popularity of cyberspace, more and more music listeners share their music preference online. Their user profiles and playlists are stored and managed in the online music databases, which are open to all the Internet users. Some popular online music applications, *e.g.*, *playlist.com*, provide services of creating user profiles and playlists, and sharing them on social networks.

MIDI Files: MIDI, an abbreviation for musical instrument digital interface, is a criterion adopted by the electronic music industry for controlling devices, such as synthesizers and sound cards, that emit music. At minimum, a MIDI representation of a sound includes values for the sound's pitch, length, and volume. It can also include additional characteristics, such as attack and delay time. The MIDI standard is supported by most synthesizers, so sounds created on one synthesizer can be played and manipulated on another synthesizer. Some free MIDI file databases provide online MIDI searching services, *e.g.*, *MIDIDB* and *IFNIMIDI*.

Music Scores: Music score refers to a hand-written or printed form of musical notation, which uses a five-line staff to represent a piece of music work. The music scores are used in playing music pieces, for example, when a pianist plays a famous piano music. In the field of music data mining, some researchers focus on music score matching, score following and score alignment, to estimate the correspondence between audio data and symbolic score [Dannenberg and Raphael, 2006]. Some popular music score websites, *e.g.*, *music-scores.com*, provide music score downloading services.

These different types of data sources represent different characteristics of music data. Music data mining aims to discover useful information and inherent features of these data sources by taking advantage of various data mining techniques. In the following, we first give a brief introduction to traditional data mining tasks, and subsequently present music related data mining tasks.

1.2 An Introduction to Data Mining

Data mining is the nontrivial extraction of implicit, previously unknown, and potentially useful information from large collection of data. The data mining process usually consists of an iterative sequence of the following steps: ***data management***, ***data preprocessing***, ***mining***, and ***post-processing*** [Li et al., 2002]. The four-component framework provides us with a simple systematic language for understanding the data mining process.

Data management is closely related to the implementation of data mining systems. Although many research papers do not explicitely laborate data management, it should note that data management can be extremely important in practical implementations. Data preprocessing is an important step to ensure the data format and quality as well as to improve the efficiency and ease of the mining process. For music data mining, especially when dealing with acoustic signals, feature extraction where the numeric features are extracted from the signals plays a critical role in the mining process. In the mining step, various data mining algorithms are applied to perform the data mining tasks. There are many different data mining tasks such as data visualization, association mining, classification, clustering, and similarity search. Various algorithms have been proposed to carry out these tasks. Finally, post-processing step is needed to refine and evaluate the knowledge derived from the mining step. Since post-processing mainly concerns the non-technical work such as documentation and evaluation, we then focus our attentions on the first three components and will briefly review data mining in these components.

1.2.1 Data Management

Data management concerns the specific mechanism and structures of how the data are accessed, stored and managed. In music data mining, data management focuses on music data quality management, involving data cleansing, data integration, data indexing and so forth.

Data Cleansing: Data cleansing refers to"cleaning" the data by filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies [Han and Kamber, 2006]. For example, in music databases, the "artists" value might be missing; we might need to set a default value for the missing data for further analysis.

Data Integration: Data integration is the procedure of combining data obtained from different data sources and providing users with an integrated and unified view of such data [Lenzerini, 2002]. This process plays a significant role in music data, *e.g.*, when performing genre classification using both acoustic features and lyrics data.

Data Indexing: Data indexing refers to the problem of storing and arranging a database of objects so that they can be efficiently searched for on the basis of their content. Particularly for music data, data indexing aims at facilitating efficient content music management [Ciaccia, 2009]. Due to the very nature of music data, indexing solutions are needed to efficiently support similarity search, where the similarity of two objects is usually defined by some expert of the domain and can vary depending on the specific application. Peculiar features of music data indexing are the intrinsic high-dimensional nature of the data to be organized, and the complexity of similarity criteria that are used to compare objects.

1.2.2 Data Preprocessing

Data preprocessing describes any type of processing performed on raw data to prepare it for another processing procedure. Commonly used as a preliminary data mining practice, data preprocessing transforms the data into a format that will be more easily and effectively processed. Data preprocessing includes data sampling, dimensionality reduction, feature extraction, feature selection, discretization, transformation and so forth.

Data Sampling: Data sampling can be regarded as a data reduction technique since it allows a large data set to be represented by a much smaller random sample (or subset) of the data [Han and Kamber, 2006]. An advantage of sampling for data reduction is that the cost of obtaining a sample is proportional to the size of the sample. Hence, sampling complexity is potentially sub-linear to the size of the data. For acoustic data, data sampling refers to measuring the audio signals at a finite set of discrete times, since a digital system such as a computer cannot directly represent a continuous audio signal.

Dimensionality Reduction: Dimensionality reduction is an important step in data mining since many types of data analysis become significantly harder as the dimensionality of the data increases, which is known as the *curse of dimensionality*. Dimensionality reduction can eliminate irrelevant features and reduce noise, which leads to a more understandable model involving fewer attributes. In addition, dimensionality reduction may allow the data to be more easily visualized. The reduction of dimensionality by selecting attributes that are a subset of the old is know as *feature selection*, which will be discussed below. Some of the most common approaches for dimensionality reduction, particularly for continues data, use techniques from linear algebra to project the data from a high-dimensional space into a lower-dimensional space, *e.g.*, *Principal Com-*

ponents Analysis (PCA) [Tan et al., 2006].

Feature Extraction: Feature extraction refers to simplifying the amount of resources required to describe a large set of data accurately. For music data, feature extraction involves low-level musical feature extraction (*e.g.*, acoustic features) and high-level features of musical feature extraction(*e.g.*, music keys). An overview of feature extraction problems and techniques is given in **Chapter ? (TO BE ADDED...)**.

Feature Selection: The purpose of feature selection is to reduce the data set size by removing irrelevant or redundant attributes (or dimensions). The goal of feature selection is to find a minimum set of attributes such that the resulting probability distribution of the data classes is as close as possible to the original distribution obtained using all features [Han and Kamber, 2006]. Feature selection can significantly improve the comprehensibility of the resulting classifier models and often build a model that generalizes better to unseen points. Further, it is often the case that finding the correct subset of predictive features is an important issue in its own right. In music data mining, feature selection is integrated with feature extraction in terms of selecting appropriate feature for further analysis.

Discretization: Discretization is used to reduce the number of values for a given continuous attributes by dividing the range of the attribute into intervals. As with feature selection, discretization is performed in a way that satisfies a criterion that is thought to have a relationship to good performance for the data mining task being considered. Typically, discretization is applied to attributes that are used in classification or association analysis [Tan et al., 2006]. In music data mining, discretization refers to breaking the music pieces down into relatively simpler and smaller parts, and the way these parts fit together and interact with each other is then examined.

Transformation: Variable transformation refers to a transformation that is applied to all the values of a variable. In other words, for each object, the transformation is applied to the value of the variable for that object. For example, if only the magnitude of a variable is important, then the values of the variable can be transformed by taking the absolute value [Tan et al., 2006]. For acoustic data, a transformation consists of any operations or processes that might be applied to a musical variable (usually a set or tone row in twelve tone music, or a melody or chord progression in tonal music) in composition, performance, or analysis. For example, we can utilize *fast fourier transform* or *wavelet transform* to transform continuous acoustic data to discrete frequency representation.

1.2.3 Data Mining Tasks and Algorithms

The cycle of data and knowledge mining comprises various analysis steps, each step focusing on a different aspect or task. Traditionally, data mining tasks involve data visualization, association mining, sequence mining, classification, clustering, similarity search, and so forth. In the following, we will briefly describe these tasks along with the techniques used to tackle these tasks.

Data Visualization

Data visualization is a fundamental and effective approach for displaying information in a graphic, tabular or other visual format [Tan et al., 2006]. The goal of visualization is to provide visual interpretations for the information being considered, and therefore the analysts can easily capture the relationship between data or the tendency of the data evolution. Successful visualization requires that the data (information) be converted into a visual format so that the characteristics of the data and the relationships among data items or attributes can be analyzed or reported. For music data, visual techniques, *e.g.*, graphs, tables and wave patterns, are often the preferred format used to explain the music social networks, music metadata and acoustic properties.

Association Mining

Association mining, the task of detecting correlations among different items in a dataset, has received considerable attention in the last few decades, particularly since the publication of the AIS and Apriori algorithms [Agrawal et al., 1993, Agrawal and Srikant, 1994]. Initially researchers on association mining were largely motivated by the analysis of market basket data, the results of which allowed companies and merchants to more fully understand customer purchasing behavior and as a result, better re-scale the market quotient. For instance, an insurance company, by finding a strong correlation between two policies A and B, of the form $A \Rightarrow B$, indicating that customers that held policy A were also likely to hold policy B, could more efficiently target the marketing of policy B through marketing to those clients that held policy A but not B. In effect, the rule represents knowledge about purchasing behavior [Ceglar and Roddick, 2006]. Another example is to find music songs patterns. Many music fans have their own playlists, in which music songs they are interested in are organized by personalized patterns. Music recommendation can be achieved by mining association patterns based on song co-occurrence.

Sequence Mining

Sequence mining is the task to find patterns which are presented in a certain number of data instances. The instances consist of sequences of elements. The detected patterns are expressed in terms of sub-sequences of the data sequences and impose an order, *i.e.*, the order of the elements of the pattern should be respected in all instances where it appears. The pattern is considered to be frequent if it appears in a number of instances above a given threshold value, usually defined by the user [Ferreira, 2005].

These patterns may represent valuable information, for example, about the customers behavior when analyzing super market transactions, or how a website should be prepared when analyzing the website log files, or when analyzing genomic or proteomic data in order to find frequent patterns which can provide some biological insights [Ferreira, 2005]. For symbolic data, a typical example of sequence mining is to recognize complex chord from MIDI guitar sequences [Scholz and Ramalho, 2008].

Classification

Classification, which is the task of assigning objects to one of several predefined categories, is a pervasive problem that encompasses many diverse applications. Examples include detecting spam email messages based upon the message header and content, classifying songs into different music genres based on acoustic features or some other music information, and categorizing galaxies based on their shapes [Tan et al., 2006]. For music data, typical classification tasks include music genre classification, artist/singer classification, mood detection, instrument recognition and so forth.

A classification technique (or called *classifier*) is a systematic approach to building classification models from an input data set. Common techniques include decision tree classifiers, rule-based classifiers, neural networks, support vector machines, and naïve Bayes classifiers [Kononenko, 1991]. Each of these techniques employs a specific learning algorithm to identify a classification model that best fits the relationship between the attribute set and class label of the input data. The model generated by a learning algorithm should both fit the input data well and correctly predict the class labels of records it has never seen before. Therefore, a key objective of the learning algorithm is to build models with good generalization capability, *i.e.*, models that accurately predict the class labels of previously unknown records [Tan et al., 2006].

Clustering

The problem of clustering data arises in many disciplines and has a wide range of applications. Intuitively, clustering is the problem of partitioning a finite set of points in a multi-dimensional space into classes (called clusters) so that (*i*) the points belonging to the same class are similar and (*ii*) the points belonging to different classes are dissimilar. The clustering problem has been studied extensively in machine learning, databases, and statistics from various perspectives and with various approaches and focuses [Li, 2005]. In music data mining, clustering involves building clusters of music tracks in a collection of popular music, identifying groups of users with different music interests, constructing music tag hierarchy and so forth.

Similarity Search

Similarity search is an important technique in a broad range of applications. To capture the similarity of complex domain-specific objects, the feature extraction is typically applied. The feature extraction aims at transforming characteristic object properties into feature values. Examples of such properties are the position and velocity of a spatial object, relationships between points on the face of a person such as the eyes, the nose, the mouth etc. The extracted values of features can be interpreted as a vector in a multidimensional vector space. This vector space is usually denoted as feature space [Pryakhin, 2006]. The most important characteristic of a feature space is that whenever two of the complex, application-specific objects are similar, the associated feature vectors have a small distance according to an appropriate distance function (e.g., the Euclidean distance). In other words, two similar, domain-specific objects should be transformed to two feature vectors that are close to each other with respect to the appropriate distance function. In contrast to similar objects, the feature vectors of dissimilar objects should be far away from each other. Thus, the similarity search is naturally translated into a neighborhood query in the feature space [Pryakhin, 2006].

Similarity search is a typical task in music information retrieval. Searching for a musical work given an approximate description of one or more of other music works is the prototype task for a music search system, and in fact it is simply addressed as similarity search. Later we will briefly introduce the task of similarity search in music information retrieval.

1.3 Music Data Mining

Music plays an important role in the everyday life for many people, and with the digitalization, large music data collections are formed and tend to be accumulated further by music enthusiasts. This has lead to music collections - not on the shelf in form of audio or video records and CD's - but on the hard drive and on the internet, to grow beyond what previously was physical possible. It has become impossible for humans to keep track of music and the relations between songs and pieces, and this fact naturally calls for data mining and machine learning techniques to assist in the navigation within the music world [Jensen, 2008]. The objective of this chapter is to explore different music data mining tasks and navigate various data mining approaches with respect to such tasks. In the following, we will introduce typical music data mining tasks, as well as the techniques employed when performing these tasks. A brief overview of these tasks and representative publications is described in Table 1.3.

1.3.1 Overview

Data mining cares about two important issues : on what kind of data and on what kind of tasks. Likewise, music data mining tries to extracting interesting patterns from various types of music data sources.

⋆ *On What Kind of Data*?

In case of a music data collection, the data sources can be many things, as we described in Table 1.1. For example, it can be the music itself, or metadata such as the title of the songs, or some acoustic features underlying audio pieces, or even statistics of how many people have listened to a track. Different analysis and experiments are conducted on such data representations based on various music data mining tasks.

⋆ *On What Kind of Tasks*?

Music data mining involves methods for various tasks, *e.g.*, genre classification, artist/singer identification, mode/emotion detection, instrument recognition, music similarity search, music summarization and visualization, and so forth. Different music data mining tasks focus on different data sources, and try to explore different aspects of data sources. For example, music genre classification aims at automatically classifying music signals into a single unique class by taking advantage of computational analysis of music feature representations [Li and Ogihara, 2005]; mood detection tries to identify the artist's emotion by virtue of acoustic features or other aspects of music data.

1.3.2 Music Data Management

Since the introduction of multimedia in personal computers, it has become more common every day to digitize part of the multimedia data, for example, music data, around us. A major advantage of digitized data over shoeboxes is that digitized data can be shared easily with others. People now create their own homepages on the world wide web, partially as a tool to manage the information they collect. But, browsing the web makes clear that a computer with a web server is not the best tool to share your "shoebox data". It is not easy for others to find your data, and the information pointed at by search engines is often incorrect, or has been moved to another location. A better solution to create large collections of music data is to organize the data in (multimedia) digital libraries [de Vries et al., 1999], *e.g.*, music databases. A digital library supports effective interaction among knowledge producers, librarians, and information and knowledge seekers. The subsequent problem of a digital library is how to efficiently store and arranged music data records so that music fans can quickly find the music resources they are interested in.

Music Indexing: A challenge in music data management is how to utilize data indexing techniques based on different aspects of the data itself. For music data, content and various acoustic features can be applied to facilitate efficient music management. For example, Shen et al. present a novel approach for generating small but comprehensive music descriptors to provide services of efficient content music data accessing and retrieval [Shen et al., 2006b]. Unlike approaches that rely on low-level spectral features adapted from speech analysis technology, their approach integrates human music perception to enhance the accuracy of the retrieval and classification process via PCA and neural networks. There are other techniques focusing on indexing music data. For instance, Crampes et al. present an innovative integrated visual approach for indexing music and for automatically composing personalized playlists for radios or chain stores [Crampes et al., 2006]. Specifically, they index music titles

Data Mining Tasks	Music Applications	Papers
Data Management	Music Database Indexing	[Crampes et al., 2006] [Rauber et al., 2002a] [Shen et al., 2006b]
Association Mining	Music Association Mining	[Knopke, 2004] [Kuo et al., 2005] [Liao et al., 2009] [Xiao et al., 2008]
Sequence Mining	Music Sequence Mining	[Bello, 2007] [Clarisse et al., 2002] [Gillet and Richard, 2007] [Guo and Siegelmann, 2004] [Peeters, 2007] [Scholz and Ramalho, 2008]
Classification	Audio Classification	[Foote, 1997] [Wold et al., 1996] [Zhang et al., 1998]
	Genre/Type Classification	[Anglade et al., 2009] [DeCoro et al., 2007] [Grimaldi et al., 2003] [Lampropoulos et al., 2005] [Li et al., 2003] [Li and Sleep, 2005] [Li and Ogihara, 2005] [Lukashevich et al., 2009] [McKay and Fujinaga, 2004] [Meng et al., 2005] [Meng and Shawe-Taylor, 2005] [Norowi et al., 2005] [Reed and Lee, 2006] [Tzanetakis and Cook, 2002] [Wang et al., 2009]
	Artist Classification	[Berenzweig et al., 2002] [Knees et al., 2004]
	Singer Identification	[Fujihara et al., 2005] [Kim and Whitman, 2002] [Liu and Huang, 2002] [Shen et al., 2006a] [Tsai et al., 2003] [Zhang, 2003]
	Mood Detection	[Hu et al., 2009] [Krumhansl, 2002] [Li and Ogihara, 2003] [Liu et al., 2003] [Tsoumakas et al., 2008] [Yang et al., 2006]
	Instrument Recognition	[Bell and Sejnowski, 1995] [Casey and Westner, 2000] [Eronen and Klapuri, 2000] [Essid et al., 2006] [Herrera-Boyer et al., 2003] [Kostek, 2004] [Sandvold et al., 2004]
Clustering	Clustering	[Camacho, 2008] [Cilibrasi et al., 2004] [Jehan, 2004] [Kameoka et al., 2005] [Li et al., 2006] [Liu et al., 2008] [Peng et al., 2007] [Pienimäki and Lemström, 2004] [Tsai et al., 2004]
Similarity Search	Similarity Search	[Aucouturier and Pachet, 2002] [Berenzweig et al., 2004] [Cooper and Foote, 2002] [Deliège et al., 2008] [Foote et al., 2002] [Li and Ogihara, 2006] [Logan and Salomon, 2001] [Nam and Berger, 2001] [Pampalk et al., 2005] [Rauber et al., 2002b] [Schnitzer et al., 2009] [Slaney et al., 2008]
Summarization	Music Summarization	[Cooper and Foote, 2002] [Cooper and Foote, 2003] [Kraft et al., 2001] [Logan and Chu, 2000] [Peeters et al., 2002] [Shao et al., 2004] [Xu et al., 2005] [Xu et al., 2004]
Data Visualization	Single Music Visualization	[Adli et al., 2010] [Chen et al., 2008] [Foote, 1999] [Foote and Cooper, 2001] [Isaacson, 2005] [Sagayama et al., 2004]
	Multi Music Visualization	[Brazil et al., 2003] [Lübbers, 2005] [Pampalk et al., 2002] [Torrens et al., 2004] [Tzanetakis and Cook, 2001]

Table 1.3: Music Data Mining Tasks.

with artistic criteria based on visual cues, and propose an integrated visual dynamic environment to assist the user when indexing music titles and editing the resulting playlists. In [Rauber et al., 2002a], Rauber et al. propose a system that automatically organizes a music collection according to the perceived sound similarity resembling genres or styles of music. In their approach, audio signals are processed according to psychoacoustic models to obtain a time-invariant representation of its characteristics. Subsequent clustering provides an intuitive interface where similar pieces of music are grouped together on a map display.

1.3.3 Music Visualization

Nowadays, people tend to enjoy the visual shock in addition to the hearing shock. As a result, more and more media player softwares provide features that generate animated imagery based on a piece of recorded music. The imagery is usually generated and rendered in real time and synchronized with the music as it is played. Actually, music data visualization is not limited to only visualizing single music document as animated imagery. The approaches to the visualization of music collections can be divided into two categories: (*i*)the ones focus on visualizing the meta or acoustic content of single music documents; (*ii*)the ones aim at representing a complete music collections for the purpose of showing the correlations among different music pieces, or grouping music pieces into different clusters based on their pairwise similarities. The former is motivated by the requirement of a casual user, when the user skims through a music CD recording before listening them carefully in order to roughly capture the main idea or the music style of the music documents. The latter is motivated by the fact that a spatial organization of the music collection will help the users finding particular songs they are interested in, because they can remember their position in the visual representation and they can be aided by the presence of similar songs nearby the searched one.

Individual Music Visualization: There are various approaches to single music visualization, most of which take advantage of music acoustic features for representing music records. For example, [Adli et al., 2010] states that symbolic analysis of music in MIDI can provide more accurate information about the musical aspects like tonality, melody line and rhythm with less computational requirements if compared with the analysis in audio files, and visualizations based on MIDI files can create visual patterns closely related to musical context as the musical information can be explicitly or implicitly obtained. Chen et al. [Chen et al., 2008] propose an emotion-based music player which synchronizes visualization (photos) with music based on the emotions evoked by auditory stimulus of music and visual content of visualization. Another example of music visualization for single music record is the pianoroll view [Sagayama et al., 2004], which proposes a new signal processing technique that provides piano-roll-like display of given polyphonic music signal with a simple transform in specmurt domain. There are some other approaches, such as the self-similarity [Foote, 1999], the plot of the waveform [Foote and Cooper, 2001], and the spectrogram [Isaacson, 2005]. Any representation has positive aspects and drawbacks, depending on the dimensions carried by the music form it is related to, and on the ability to capture relevant features. Representations can be oriented toward a global representation or local characteristics.

Music Collection Visualization: Visualization of a collection of musical records is usually based on the concept of similarity. Actually, the problem of a graphical representation, normally based on bidimensional plots, is typical of many areas of data analysis. Techniques such as *Multidimensional Scaling* and *Principal Component Analysis* are well-known for representing a complex and multidimensional set of data when a distance measure – such as the musical similarity – can be computed between the elements or when the elements are mapped to points in a high-dimensional space. The application of bidimensional visualization techniques to music collections has to be carried out considering that the visualization will be given to non-expert users, rather than to data analysts, who need a simple and appealing representation of the data.

One example of system for graphical representation of audio collection is Marsyas3D [Tzanetakis and Cook, 2001], which includes a variety of alternative 2D and 3D representations of elements in the collection. In particular, *Principal Component Analysis* is used to reduce the parameter space that describes the timbre in order to obtain either a bidimensional or tridimensional representation. Another example is the Sonic Browser, which is an application for browsing audio collections [Brazil et al., 2003] that provides the user with multiple views, including a bidimensional scatterplot of audio objects, where the coordinates of each point depend on attributes of the dataset, and a graphical tree representation, where the tree is depicted with the root at the center and the leaves over a circle. The sonic radar, presented in [Lübbers, 2005], is based on the idea that only a few objects, called prototype songs, can be presented to the user. Each prototype song is obtained through clustering the collection with k-means algorithm and extracting the song that is closer to the cluster center. Prototype songs are plotted on a circle around a standpoint. In addition, Torrens et al. propose different graphical visualization views and their associated features to allow users to better organize their personal music libraries and therefore also to ease selection later on [Torrens et al., 2004]. [Pampalk et al., 2002] presents a system with islands of music which facilitates exploration of music libraries without requiring manual genre classification. Given pieces of music in raw audio format they estimate their perceived sound similarities based on psychoacoustic models. Subsequently, the pieces are organized on a 2-dimensional map so that similar pieces are located close to each other.

1.3.4 Music Information Retrieval

Music Information Retrieval (MIR) is an emerging research area devoted to fulfill users' music information needs. As it will be seen, despite the emphasis on retrieval of its name, MIR encompasses a number of different approaches aiming at music management, easy access, and enjoyment. Most of the research work on MIR, of the proposed techniques, and of the developed systems are content based [Orio, 2006]. The main idea underlying content-based approaches is that a document can be described by a set of features that are directly computed from its content. In general, content-based access to multimedia data requires specific methodologies that have to be tailored to each particular medium. Yet, the core information retrieval (IR) techniques, which are based on statistics and probability theories, may be more generally employed outside the textual case, because the underlying models are likely to describe fundamental characteristics being shared by different media, languages, and application domains [Jones and Willett, 1997].

A great variety of different methods for content-based searching in music scores and audio data have been proposed and implemented in research prototypes and commercial systems. Besides the limited and well-defined task of identifying recordings, for which audio fingerprinting techniques work well, it is hard to tell which methods should be further pursued. This underlines the importance of a TREC-like series of comparisons for algorithms (such as EvalFest/MIREX at ISMIR) for searching audio recordings and symbolic music notation. Audio and symbolic methods are useful for different tasks. For example, identification of instances of recordings must be based on audio data, while works are best identified based on a symbolic representation. For determining the genre of a given piece of music, approaches based on audio look promising, but symbolic methods might work as well. The interested reader can get a brief overview of different content-based music information retrieval systems in [Typke et al., 2005].

Music Similarity Search: In the field of music data mining, similarity search refers to searching for music sound files similar to a given music sound file. In principle, searching can be carried out on any dimension. For instance, the user could provide an example of the timbre – or of the sound – that he is looking for, or describe the particular structure of a song, and then the music search system will be search similar music works based the information given by the user.

The similarity search processes can be divided into feature extraction and query processing [Li and Ogihara, 2006]. For feature extraction, detailed procedure or instruction is introduced in **Chapter ? TO BE ADDED...**. After feature extraction, music works can be represented based on the extracted features. In the step of query processing, the main task is to employ a proper similarity measure to calculate the similarity between the given music work and the candidate music works. A variety of existing similarity measures and distance functions have previously been examined in this context, spanning from simple Euclidean and Mahalanobis distances in feature space to information theoretic measures like the Earth Mover Distance and Kullback-Leibler [Berenzweig et al., 2004]. Regardless of the final measure, a major trend in the music retrieval community has been to use a density model of the features (often timbre space defined by MFCC's [Rabiner and Juang, 1993]). The main task of comparing two models has then been handled in different ways and is obviously the more interesting task.

The objective of similarity search is to find music sound files similar to a music sound file given as input. Music classification based on genre and style is naturally the form of a hierarchy. Similarity can be used to group sounds together at any node in the hierarchies. The use of sound signals for similarity is justified by an observation that audio signals (digital or analog) of music belonging to the same genre share certain characteristics, because they are composed of similar types of instruments, having similar rhythmic patterns, and similar pitch distributions [Dowling and Harwood, 1986].

The problem of finding sound files similar to a given sound files has been studied in the last decade. Logan and Salomon propose the use of MFCC to define similarity [Logan and Salomon, 2001]. Nam and Berger propose the use of timbral features (spectral centroids, short-term energy function, and zero-crossing) for similarity testing [Nam and Berger, 2001]. Cooper and Foote study the use of self-similarity to summary music signals [Cooper and Foote, 2002]. And then in the subsequent work they use this summarization for retrieving music files [Foote et al., 2002]. Rauber, Pampalk, and Merkl study a hierarchical approach in retrieving similar music sounds [Rauber et al., 2002b]. Schnitzer et al. rescale the divergence and use a modified FastMap implementation to accelerate nearest-neighbor queries [Schnitzer et al., 2009]. Slaney et al. learn embeddings so that the pairwise Euclidean distance between two songs reflects semantic dissimilarity [Slaney et al., 2008]. Deliège et al. perform the feature extraction in a two-step process that allows distributed computations while respecting copyright laws [Deliège et al., 2008]. Li and Ogihara investigate the use of acoustic based features for music information retrieval [Li and Ogihara, 2006]. For similarity search, the distance between two sound files is defined to be the Euclidean distance of their normalized representations. Pampalk et al. present an approach to improve audio-based music similarity and genre classification [Pampalk et al., 2005]. Berenzweig et al. examine both acoustic and subjective approaches for calculating similarity between artists, comparing their performance on a common database of 400 popular artists [Berenzweig et al., 2004]. Aucouturier and Pachet introduce a timbral similarity measures for comparing music titles based on a Gaussian model of cepstrum coefficients [Aucouturier and Pachet, 2002].

1.3.5 Association Mining

As discussed in Section 1.2.3, association mining refers to detecting correlations among different items in a dataset. Specifically in music data mining, association mining can be divided into three different categories: (*i*) detecting associations among different acoustic features, for example, Xiao et al. use statistic model to investigate the association between timbre and tempo and use timbre information to improve the performance of tempo estimation [Xiao et al., 2008]; (*ii*) detecting associations among music and other document formats. For instance, [Knopke, 2004] measures the similarity between the public text visible on a web page and the linked sound files, the name of which is normally unseen by the user. Liao et al. use a dual-wing harmonium model to learn and represent the underlying association patterns between music and video clips in professional MTV [Liao et al., 2009]; (*iii*) detecting associations among music features and other music aspects, *e.g.*, emotions. An illustrative example of research papers related this category is [Kuo et al., 2005], which investigates the music feature extraction and modified the affinity graph for association discovery between emotions and music features. Such affinity graph can provide insight for music recommendation.

1.3.6 Sequence Mining

For music data, sequence mining mainly aims to detect sub-sequences from different audio aspects, for example, chord sequences. There are relatively few publications related to music sequence mining tasks. The main contribution of sequence mining is in the field of music transcription. When transcribing audio pieces, different types of errors might be introduced in the transcription version, such as segmentation errors, substitution errors, time alignment errors and so on. To better control the error rate of the transcription, researchers try to explore the feasibility of applying sequence mining into the procedure of transcribing music recordings. For example, Gillet and Richard discuss two post-processings for drum transcription systems, which aim to model typical properties of drum sequences [Gillet and Richard, 2007]. Both methods operate on a symbolic representation of the sequence, which is obtained by quantizing the onsets of drum strokes on an optimal tatum grid, and by fusing the posterior probabilities produced by the drum transcription system. In [Clarisse et al., 2002] a new system is presented for the automatic transcription of singing sequences into a sequence of pitch and duration pairs. Guo et al. present the Time-Warped Longest Common Subsequence algorithm (T-WLCS), which deals with singing errors involving rhythmic distortions [Guo and Siegelmann, 2004]. Other applications of sequence mining for music data include chord sequence detection [Bello, 2007, Scholz and Ramalho, 2008] and exploring music structure [Peeters, 2007].

1.3.7 Classification

Classification is an important issue within music data mining tasks. Various classification problems emerge in the recent decades. Some researchers focus on classifying music from audio pieces, whereas others are engaged in categorizing music works into different groups. The most general classification focuses on music genre/style classification. In addition, there are some other classification tasks, such as artist/singer classification, mood/emoteion classification, instrument classification and so on. In the following, we will provide a brief overview on different classification tasks. Table 1.4 briefly summarizes different classification tasks.

Audio Classification: The term audio classification has been traditionally used to describe a particular task in the fields of speech and video processing, where the main goal is to identify and label the audio in three different classes: speech, music, and environmental sound. This first coarse classification can be used to aid video segmentation or decide where to apply automatic speech recognition. The refinement of the classification with a second step, where music

Tasks	Techniques	Papers
Audio Classification	Tree-based Quantization	[Foote, 1997]
	Covariance Matrix	[Wold et al., 1996]
	Hidden Markov Model	[Zhang et al., 1998]
Genre Classification	Bayesian Model	[DeCoro et al., 2007]
	Decision Tree	[Anglade et al., 2009]
	Hidden Markov Model	[Reed and Lee, 2006]
	Statistical Pattern Recognition	[Tzanetakis and Cook, 2002]
	Wavelet Transformation	[Grimaldi et al., 2003] [Li et al., 2003]
	SVM	[Li and Sleep, 2005] [Meng and Shawe-Taylor, 2005]
	Taxonomy	[Li and Ogihara, 2005]
	Multi-labeling Classification	[Lukashevich et al., 2009] [Wang et al., 2009]
	Neural Networks	[Lampropoulos et al., 2005] [McKay and Fujinaga, 2004] [Meng et al., 2005]
Artist Classification	Singer Voice	[Berenzweig et al., 2002]
	Text Categorization	[Knees et al., 2004]
Singer Identification	Gaussian Mixture Model	[Fujihara et al., 2005] [Kim and Whitman, 2002] [Shen et al., 2006a] [Tsai et al., 2003] [Zhang, 2003]
	KNN	[Liu and Huang, 2002]
Mood Detection	SVM on Text Features	[Hu et al., 2009]
	Multi-label Classification	[Li and Ogihara, 2003] [Tsoumakas et al., 2008]
	Fuzzy Classifier	[Yang et al., 2006]
	Gaussian Mixture Model	[Liu et al., 2003]
Instrument Recognition	Statistical Model	[Eronen and Klapuri, 2000] [Essid et al., 2006]
	Neural Networks	[Kostek, 2004]
	Prior Knowledge	[Sandvold et al., 2004]
	Taxonomy	[Herrera-Boyer et al., 2003]

Table 1.4: Music Classification.

signals are labeled with a number of predefined classes, has been presented in [Zhang et al., 1998], which is also worth mentioning because it is one of the first papers that present hidden Markov models as a tool for MIR. An early work on audio classification, presented in [Wold et al., 1996], was aimed at retrieving simple music signals using a set of semantic labels, in particular focusing on the musical instruments that are part of the orchestration. The approach is based on the combination of segmentation techniques with automatic separation of different sources and the parameter extraction. The classification based on the particular orchestration is still an open problem with complex polyphonic performances.

An important issue in audio classification, introduced in [Foote, 1997], is the amount of audio data needed to achieve good classification rates. This problem has many aspects. First, the amount of data needed is strictly related to the computational complexity of the algorithms, which usually are at least linear with the number of audio samples. Second, perceptual studies showed that even untrained listeners are quite good at classifying audio data with very short excerpts (less than 1 sec). Finally, in a query-by-example paradigm, where the examples have to be digitally recorded by users, it is likely that users will not be able to record a large part of audio.

Genre Classification: A particular aspect of music record classification is genre classification. The problem is to correctly label an unknown recording of a song with a music genre. Labels can be hierarchically organized in the collection of genres and subgenres. Labeling can be used to enrich the musical document with high-level metadata or to organize a music collection. Genre classification is still biased by Western music, and thus genres are the ones typically found in Western music stores. Some attempts have been made to extend the approach to other cultures, for instance in [Norowi et al., 2005], genre classification has been carried for traditional Indian musical forms together with Western genres.

One of the first papers introducing the problem of music classification is [Tzanetakis and Cook, 2002], in which there is a bias toward classical music and jazz, while some genresambient, electronic, and ethnicare not reported. This is a typical problem of music classification, because the relevance of the different categories is extremely subjective, as well as the categories themselves. These problems are faced also by human classifiers that try to accomplish the same task, and in fact in [Tzanetakis and Cook, 2002] it is reported that college students achieved no more than about 70% of classification accuracy when listening to three seconds of audio (listening to longer excerpt did not improve the performances). The automatic classification is based on three different feature sets, related to rhythmic, pitch, and timbre features. As also highlighted in subsequent works, rhythm seems to play an important role for the classification.

The features used as content descriptors are normally the ones related to timbre. This choice depends on the fact that approaches try to classify short excerpts of an audio recording, where middle-term features like melody and harmony are not captured. Common music processing approaches compute the *Mel-Frequency Cepstral Coefficients* (MFCCs), while the use of the wavelet transform is exploited in [Grimaldi et al., 2003] and [Li et al., 2003]. Systems on genre classification are normally trained with a set of labeled audio excerpts, and classification is carried out using different techniques and models from the classification literature. In particular, different feature extractions [Anglade et al., 2009, Lampropoulos et al., 2005, Li et al., 2003, McKay and Fujinaga, 2004, Meng et al., 2005, Reed and Lee, 2006, Tzanetakis and Cook, 2002] and Gaussian Mixtures Models [Meng and Shawe-Taylor, 2005] are classically used for classifying genres, but Support Vector Machines [Li and Sleep, 2005, Wang et al., 2009] and Bayesian theories [DeCoro et al., 2007] have also been successfully applied to this task. Also, some other aspects of music, such as taxonomies [Li and Ogihara, 2005], can be applied to tackling genre classification problem. Advanced data mining techniques, such as multi-label classification [Lukashevich et al., 2009, Wang et al., 2009], provide efficient approaches for this problem.

Mood Detection and Classification: Music mood describes the inherent emotional meaning of a music clip. It is helpful in music understanding, music search and some music-related applications. One common opinion objecting to mood detection is that the emotional meaning of music is subjective and it depends on many factors including culture. Music psychologists now agree that culture is of great importance in people's mood response to music, as well as other factors including education and previous experiences. Krumhansl [Krumhansl, 2002] also pointed out that musical sounds might inherently have emotional meaning. For example, some music patterns represent contentment or relaxing, while some others make an individual feel anxious or frantic.

Recently, there are some research works touching the field of music mood detection and classification. Liu et al. present a hierarchical framework to automate the task of mood detection from acoustic music data, by following some music psychological theories in western cultures [Liu et al., 2003]. Yang et al. consider a different approach to music emotion classification [Yang et al., 2006]. For each music segment, the approach determines how likely the song segment belongs to an emotion class. Two fuzzy classifiers are adopted to provide the measurement of the emotion strength. Hu et al. investigate the usefulness of text features in music mood classification on 18 mood categories derived from user tags [Hu et al., 2009].

In addition, some advanced data mining techniques are applied to music mood detection and classification, for example, multi-label classification. Li and Ogihara cast the emotion detection problem as a multi-label classification problem, where the music sounds are classified into multiple classes simultaneously [Li and Ogihara, 2003]. In [Tsoumakas et al., 2008], the automated detection of emotion in music is modeled as a multi-label classification task, where a piece of music may belong to more than one class.

Instrument Recognition and Classification: The need for automatic classification of sounds arises in different contexts, such as biology (*e.g.*, for identifying animals belong to a given species or for cataloguing communicative resources), medical diagnosis (*e.g.*, for detecting abnormal conditions of vital organs), surveillance (*e.g.*, for recognizing machine-failure conditions), military operations (*e.g.*, for detecting an enemy engine approaching or for weapon identification) and multimedia content description (*e.g.*, for helping video scene classification or instrument recognition). In this section, we focus on describing sound effects in the case of music, which means description calls for deriving indexes in order to locate melodic patterns, harmonic or rhythmic structures, etc [Herrera-Boyer et al., 2003].

Music instrument recognition and classification are very difficult tasks that are far from being solved. The practical utility for musical instrument classification is twofold:

1. to provide labels for monophonic recordings, for "sound samples" inside sample libraries, or for new patches created with a given synthesizer;
2. to provide indexes for locating the main instruments that are included in a musical mixture (for example, one might want to locate a saxophone "solo" in the middle of a song).

The first problem is easier to solve than the second one, and it seems clearly solvable given the current state of the art. The second is tougher, and it is not clear if research done on solving the first one may help. Common sense dictates that a reasonable approach to the second problem would be the initial separation of the sounds corresponding to the different sound sources, followed by the segmentation and classification on those separated tracks. Techniques for source separation cannot yet provide satisfactory solutions although some promising approaches have been developed [Bell and Sejnowski, 1995, Casey and Westner, 2000]. As a consequence, research on music instrument classification has concentrated on working with isolated sounds under the assumption that separation and segmentation have been previously performed. This implies the use of a sound sample collection (usually isolated notes) consisting of different instrument families and classes.

Most publications on music instrument recognition and classification focus on analyzing acoustic features of sounds in the music pieces. For example, Eronen et al. present a system for musical instrument recognition that takes advantage of a wide set of features to model the temporal and spectral characteristics of sounds [Eronen and Klapuri, 2000]. In [Kostek, 2004], instrument classification process is shown as a three-layer process consisting of pitch extraction, parametrization, and pattern recognition. Sandvold et al. present a feature-based sound modeling approach that combines general, prior knowledge about the sound characteristics of percussion instrument families (general models) with on-the-fly acquired knowledge of recording-specific sounds (localized models) [Sandvold et al., 2004]. Essid et al. utilize statistical pattern recognition techniques to tackle the problem in the context of solo musical phrases [Essid et al., 2006]. Readers interested in this research area can get some insight in [Herrera-Boyer et al., 2003], which presents an exhaustive review of research on automatic classification of sounds from musical instruments.

Artist Classification: The term artist classification refers to classifying musicians as the predefined artist label given a music document. Traditionally, artist classification is performed based on acoustic features or singer voice.

For instance, Berenzweig et al. present that automatically-located singing segments form a more reliable basis for classification than using the entire track, suggesting that the singer's voice is more stable across different performances, compositions, and transformations due to audio engineering techniques than the instrumental background [Berenzweig et al., 2002]. An alternative approach to artist classification is to utilize text categorization techniques to classify artists. Knees et al. retrieve and analyze web pages ranked by search engines to describe artists in terms of word occurrences on related pages [Knees et al., 2004].

Singer Identification: Automated singer identification is important in organizing, browsing and retrieving data in large music collections due to numerous potential applications including music indexing and retrieval, copyright management and music recommendation systems. The development of singer identification enables the effective management of music databases based on "singer similarity". With this technology, songs performed by a particular singer can be automatically clustered for easy management or exploration, as described in [Shen et al., 2006a].

Several approaches have been proposed to take advantage of statistical models or machine learning techniques for automatic singer classification/identification [Kim and Whitman, 2002, Liu and Huang, 2002, Tsai et al., 2003, Zhang, 2003]. In general, these methods consist of two main steps: singer characteristic modelling based on solo voice and class label identification. In the singer characteristic modelling step, acoustic signal information is extracted to represent the music. Then specific mechanisms (e.g., statistical models or machine learning algorithms) are constructed to assign the songs to one of the pre-defined singer categories based on their extracted acoustic features. In addition, Shen et al. use multiple low-level features extracted from both vocal and non-vocal music segments to enhance the identification process with a hybrid architecture and build profiles of individual singer characteristics based on statistical mixture models [Shen et al., 2006a]. Fujihara et al. describe a method for automatic singer identification from polyphonic musical audio signals including sounds of various instruments [Fujihara et al., 2005].

1.3.8 Clustering

Clustering, as introduced in Section 1.2.3, is the task of separating a collection of data into different groups based on some criteria. Specifically in music data mining, clustering aims at dividing a collection of music data into groups of similar objects based on their pairwise similarities without predefined class labels.

There are a couple of research publications related to music data clustering. For instance, in [Liu et al., 2008], an inter-recording distance metric which characterizes diversity of pitch distribution together with harmonic center of music pieces, is introduced to measure dissimilarities among musical features, based on chroma-based features extracted from acoustic signals. Camacho tracks the pitch strength trace of the signal, determining clusters of pitch and unpitched sound based on the criterion of the local maximization of the distance between the centroids [Camacho, 2008]. Tsai et al. examine the feasibility of unsupervised clustering of music data based on their singer. It has been shown that the characteristics of a singer's voice can be extracted from music via vocal segment detection followed by solo vocal signal modeling [Tsai et al., 2004]. Li et al. propose a clustering algorithm that integrates features from both lyrics and acoustic data sources to perform bimodal learning [Li et al., 2006]. In order to reduce the dimensionality of music features, Jehan propose a perceptually grounded model for describing music as a sequence of labeled sound segments, for reducing data complexity, and for compressing audio [Jehan, 2004]. In addition, some simple data representation formats are introduced when performing clustering music collections. An illustrative example is introduced in [Pienimäki and Lemström, 2004], in which Pienimäki et al. propose a novel automatic analysis method based on paradigmatic and surface level similarity of music represented in symbolic form. Kameoka et al. decompose the energy patterns diffused in time frequency space, *i.e.*, a time series of power spectrum, into distinct clusters such that each of them is originated from a single sound stream [Kameoka et al., 2005]. Peng et al. propose an approach based on the generalized constraint clustering algorithm by incorporating the constraints for grouping music by "similar" artists [Peng et al., 2007]. In [Cilibrasi et al., 2004], Cilibrasi et al. apply compression-based method to the clustering of pieces of music.

1.3.9 Music Summarization

How to create a concise and informative extraction that best summarizes an original digital content is another challenge in music data mining and is extremely important in large-scale information organization and understanding [Shao et al., 2004]. Recently, most of the music summarization for commercial use is manually generated from the original music recordings. However, since a large volume of digital content has been made publicly available on mediums, for example, the Internet, efficient approaches of automatic music summarization have become increasingly important and necessary.

Similar to text summarization, music summarization aims to determine the most general and salient themes of a given music piece that may be used as a representative of the music and readily recognized by a listener. Automatic music summarization can be applied to music indexing, content-based music retrieval and web-based music distribution [Xu et al., 2004]. Several research methods are proposed in automatic music summarization. A music summarization system [Kraft et al., 2001] was developed on MIDI format, which utilized the repetition nature of MIDI compositions to automatically recognize the main melody theme segment of a given piece of music and generate music summary. But MIDI is a synthesizer and structured format and is different from sampled audio format such as wav which is highly unstructured. Therefore, MIDI summarization method cannot be applied to real music summarization. A real music summarization system [Logan and Chu, 2000] used MFCCs to parameterize each music song. Based on MFCCs, a cross-entropy measure, or HMM, was used to discover the song structure. Then heuristics were applied to extract the key phrase in terms of this structure. This summarization method is suitable for certain genres of music such as rock or folk music, but it is less applicable to classical music. MFCCs were also used as features in Cooper and Foote's works [Cooper and Foote, 2002, Cooper and Foote, 2003]. They used a two-dimensional (2-D) similarity matrix to represent music structure and generate music summary. But this approach will not always yield intuitively piece. Peeters et al. [Peeters et al., 2002] proposed a multi-pass approach to generate music summaries. [Xu et al., 2005] proposes effective algorithms to automatically classify and summarize music content. Support vector machines are applied to classify music into pure music and vocal music by learning from training data. For pure music and vocal music, a number of features are extracted to characterize the music content, respectively. Based on calculated features, a clustering algorithm is applied to structure the music content. Finally, a music summary is created based on the clustering results and domain knowledge related to pure and vocal music.

1.3.10 Advanced Music Data Mining Tasks

In the wake of the increasing popularity of music and the avalanche of various music applications and softwares, the research directions of music data mining tend to be diverse. Specifically, advanced data mining techniques based on different learning metrics emerge in music data mining community. A couple of learning tasks, involving multi-task learning, multi-instance learning, multi-label classification and so on, are introduced in the following:

Multi-Task: Multi-task learning (MTL) [Caruana, 1997] has attracted significant interest in data mining and machine learning community over the last decade [Blei et al., 2004, Thrun and O'Sullivan, 1996, Bakker and Heskes, 2003, Xue et al., 2007]. Many of the research publications on MTL have explored ideas in Bayesian hierarchical modeling [Gelman, 2004], and such approaches have been successfully applied to information retrieval [Blei et al., 2004] and computer vision [Thrun and O'Sullivan, 1996]. For music data, multi-task learning has comprehensive applications. For example, Ni et al. [Ni et al., 2008] employ a nonparametric Bayesian approach [Teh et al., 2006] for multi-task learning in which the number of states is not fixed a priori; the model is termed an infinite HMM (iHMM). To learn multiple iHMMs simultaneously, one for each sequential data set, the base distributions of the iHMMs may be drawn from an nDP [Rodriguez et al., 2008], this allowing inter-task clustering.

Multi-Instance: Multiple-instance learning (MIL) [Dietterich et al., 1997] is active in the classification tasks. MIL algorithms train classifiers from lightly supervised data, *e.g.*, labeled collections of items, known as *bags*, rather than labeled items. Particularly, in music data mining, there are many high quality sources of metadata about musical information such as *Last.fm*, the *All Music Guide*, *Pandora.com*, etc. However, each source provides metadata only at certain granularities, *i.e.*, describes the music only at certain scales. For music data, clip(part of tracks)-level classifiers can be used to refine descriptions from one granularity to finer granularities, e.g. using audio classifiers trained on descriptions of artists to infer descriptions of albums, tracks, or clips. This metadata refinement problem is a MIL problem. Some publications are related to this research area, for instance, Mandel and Ellis [Mandel and Ellis, 2008] formulated a number of music information related multiple-instance learning tasks and evaluated the mi-SVM and MILES algorithms on them.

Multi-Label: In machine learning, multi-label classification is the special case within classification of assigning one of several class labels to an input object. Unlike the better understood problem of binary classification, which requires discerning between the two given classes, the multi-label one is a more complex and less researched problem. Recently, multi-label classification is increasingly applied in music categorization problem. For example, Wang et al. [Wang et al., 2009] propose a multi-label music style classification approach, called Hypergraph integrated Support Vector Machine (HiSVM), which can integrate both music contents and music tags for automatic music style classification. Li and Ogihara [Li and Ogihara, 2003] cast the emotion detection problem as a multi-label classification problem, where the music sounds are classified into multiple classes simultaneously.

Semi-Supervised Learning: Semi-supervised learning is a type of machine learning techniques that makes use of both labeled and unlabeled data for training – typically a small amount of labeled data with a large amount of unlabeled data. Semi-supervised learning falls between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data). Many machine-learning researchers have found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce considerable improvement in learning accuracy. The acquisition of labeled data for a learning problem often requires a skilled human agent to manually classify training examples. The cost associated with the labeling process thus may render a fully labeled training set infeasible, whereas acquisition of unlabeled data is relatively inexpensive. In such situations, semi-supervised learning can be of great practical value. In music data mining, semi-supervised learning can be applied to the task of classifying music metadata. For instance, Li and Ogihara [Li and Ogihara, 2004] study the problem of identifying "similar" artists using both lyrics and acoustic data. The approach uses a small set of labeled samples for the seed labeling to build classifiers that improve themselves using unlabeled data. You et al. [You and Dannenberg, 2007] explore the use of machine learning to improve onset detection functions. To solve the problem of training data, they use a semi-supervised learning technique combined with score alignment. The result of alignment is an estimate of the onset time of every note in the MIDI file, and these estimates are improved by iteratively applying our onset detector and then retraining on the new data.

Tensor-based Learning: A *tensor* is a multidimensional array. More formally, an N-way or Nth-order tensor is an element of the tensor product of N vector spaces, each of which has its own coordinate system. Decompositions of higher-order tensors (*e.g.*, N-way arrays with $\mathrm{N} \geq 3$) have applications in psychometrics, chemometrics, signal processing, numerical linear algebra, computer vision, numerical analysis, data mining, neuroscience, graph analysis, and elsewhere [Kolda and Bader, 2009]. Particularly, in data mining, tensor decomposition and factorization has comprehensive applications. For example, Liu et al. [Liu et al., 2005] propose a text representation model, Tensor Space Model (TSM), which models the text by multi-linear algebraic high-order tensor instead of the traditional vector. Sun et al. [Sun et al., 2006] introduce the dynamic tensor analysis (DTA) method to summarize high-order or high-dimensional data and reveal the hidden correlations among data. In music data mining, tensor analysis has its own applications in different aspects. Benetos et al. [Benetos and Kotropoulos, 2008] propose an automatic music genre classification system using tensor representations, where each recording is represented by a feature matrix over time. In [FitzGerald et al., 2006], an algorithm which performs shifted Non-negative Tensor Factorization is presented to separate harmonic instruments from multichannel recordings, extending shifted Non-negative Matrix Factorization to the multi-channel case.

1.4 Conclusions

The term music data mining encompasses a number of different research and development activities that have the common denominator of being related to music data access and analysis [Orio, 2006]. In this chapter, we have introduced the state of the art of music data mining, navigating from the basic data mining tasks to specific music data mining applications. Some popular music data mining tasks, such as music genre classification, singer identification, emotion detection, instrument recognition and so on, provide substantial benefits for the real world music management applications and softwares. In addition to the research discussed above, there are many other research issues in music data mining.

Data issues: These include mining various music data characteristics and information from heterogeneous music databases, and the use of culture knowledge.

- *Mining different kinds of features in music*: Music, to some extent, represents information combination involving cultures, artists' interests and so on. The very nature of music data may provide various unexplored features for research purpose, which denote diverse audio aspects in terms of music itself. For example, when an artist is playing a melody, the keynote (or mood) may change accompanied by the progress of the melody. By extracting related features that can detect the emotion of the artists, we can easily identify the emotion variation in this melody, which helps us to understand the artist's sentiment on the objects of reference when he/she is composing.

- *Mining information from heterogeneous music data sources*: Various types of data describing different aspects related to music emerge in the wake of the explosion of music information, such as music reviews, music tags and so on. The procedure of music data mining tasks may be facilitated by taking into account such information. For instance, by virtue of music tags, we can somehow improve music genre classification; by incorporating music reviews, music emotion detection could be more easier and reasonable. Moreover, we can explore the inner-relationship among different music data sources, *e.g.*, users, songs, artists, tags, reviews and so on, and then deduce a high level music social network.

- *Incorporation of culture knowledge*: A particular characteristics of music data is the culture difference. Music is an art form that can be shared by people with different culture because it crosses the barriers of national languages and cultural backgrounds. For example, western classical music has passionate followers in China, and many persons in Europe are keen on classical Indian music: all of them can enjoy music without the need of a translation, which is normally required for accessing foreign textual works [Orio, 2006]. Therefore, how to eliminate culture difference in a reasonable and understandable way is a special research direction.

Methodology issues:

- *Interactive mining of music information*: Interactive mining allows users to focus on the search for patterns, providing and refining data mining requests based on returned results [Han and Kamber, 2006]. It is an important aspect of music information retrieval systems, since effective interactive mining can help users to better describe their music information needs. For example, we can utilize dynamic query form to facilitate interactions between users and MIR systems.

Performance issues: These include efficiency and scalability of music data mining algorithms, and the visualization of music data mining results.

- *Efficiency and scalability of music data mining algorithms*: Music databases consist of a huge amount of data, *e.g.*, music metadata, acoustic features and so on. A natural question is how to effectively extract information from such databases. This requires music data mining algorithms being efficient and scalable [Han and Kamber, 2006]. In other words, the running time of a music data mining algorithm must be predictable and acceptable in large databases. The efficiency and scalability of algorithms are then becoming key issues in the implementation of music data mining systems.

- *Visualization of music data mining results*: Music data mining results are usually represented as tables or simple plots, which cannot be vividly analyzed. In addition, some patterns hidden in the results cannot be easily identified by reviewing these simple representations. This is especially crucial if the music data mining system is to be interactive.

The above issues are regarded as major challenges for the further evolution of music data mining technology.

Bibliography

[Adli et al., 2010] Adli, A., Nakao, Z., and Nagata, Y. (2010). A content dependent visualization system for symbolic representation of piano stream. In *Knowledge-Based Intelligent Information and Engineering Systems*, pages 287–294.

[Agrawal et al., 1993] Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 22(2):207–216.

[Agrawal and Srikant, 1994] Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules. In *Proceeding of the 20th International Conference on Very Large Data Bases, VLDB*, volume 1215, pages 487–499.

[Anglade et al., 2009] Anglade, A., Mary, Q., Ramirez, R., and Dixon, S. (2009). Genre classification using harmony rules induced from automatic chord transcriptions. In *Proceedings of the International Conference on Music Information Retrieval*, pages 669–674.

[Aucouturier and Pachet, 2002] Aucouturier, J. and Pachet, F. (2002). Music similarity measures: What's the use? In *Proceedings of the International Conference on Music Information Retrieval*, pages 157–163.

[Bakker and Heskes, 2003] Bakker, B. and Heskes, T. (2003). Task clustering and gating for bayesian multitask learning. *The Journal of Machine Learning Research*, 4:83–99.

[Bell and Sejnowski, 1995] Bell, A. and Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159.

[Bello, 2007] Bello, J. (2007). Audio-based cover song retrieval using approximate chord sequences: testing shifts, gaps, swaps and beats. In *Proceedings of the International Conference on Music Information Retrieval*, pages 239–244.

[Benetos and Kotropoulos, 2008] Benetos, E. and Kotropoulos, C. (2008). A tensor-based approach for automatic music genre classification. In *Proceeding of the 16th European Conference on Signal Processing*.

[Berenzweig et al., 2002] Berenzweig, A., Ellis, D., and Lawrence, S. (2002). Using voice segments to improve artist classification of music. In *AES 22nd International Conference*, pages 79–86.

[Berenzweig et al., 2004] Berenzweig, A., Logan, B., Ellis, D., and Whitman, B. (2004). A large-scale evaluation of acoustic and subjective music-similarity measures. *Computer Music Journal*, 28(2):63–76.

[Blei et al., 2004] Blei, D., Griffiths, T., Jordan, M., and Tenenbaum, J. (2004). Hierarchical topic models and the nested Chinese restaurant process. *Advances in neural information processing systems*, 16:17–24.

[Brazil et al., 2003] Brazil, E. et al. (2003). Audio information browsing with the sonic browser. pages 26–31.

[Camacho, 2008] Camacho, A. (2008). Detection of pitched/unpitched sound using pitch strength clustering. In *Proceedings of the International Conference on Music Information Retrieval*, pages 533–537.

[Caruana, 1997] Caruana, R. (1997). Multitask learning. *Machine Learning*, 28(1):41–75.

[Casey and Westner, 2000] Casey, M. and Westner, A. (2000). Separation of mixed audio sources by independent subspace analysis. In *Proceedings of the International Computer Music Conference*.

[Ceglar and Roddick, 2006] Ceglar, A. and Roddick, J. (2006). Association mining. *ACM Computing Surveys (CSUR)*, 38(2).

[Chen et al., 2008] Chen, C., Weng, M., Jeng, S., and Chuang, Y. (2008). Emotion-based music visualization using photos. *Advances in Multimedia Modeling*, pages 358–368.

[Ciaccia, 2009] Ciaccia, P. (2009). Multimedia Data Indexing.

[Cilibrasi et al., 2004] Cilibrasi, R., Vitányi, P., and Wolf, R. (2004). Algorithmic clustering of music based on string compression. *Computer Music Journal*, 28(4):49–67.

[Clarisse et al., 2002] Clarisse, L., Martens, J., Lesaffre, M., De Baets, B., De Meyer, H., and Leman, M. (2002). An auditory model based transcriber of singing sequences. In *Proceedings of the International Conference on Music Information Retrieval*, pages 116–123.

[Cooper and Foote, 2002] Cooper, M. and Foote, J. (2002). Automatic music summarization via similarity analysis. In *Proc. IRCAM*, pages 81–85.

[Cooper and Foote, 2003] Cooper, M. and Foote, J. (2003). Summarizing popular music via structural similarity analysis. In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop*, pages 127–130.

[Crampes et al., 2006] Crampes, M., Ranwez, S., Velickovski, F., Mooney, C., and Mille, N. (2006). An integrated visual approach for music indexing and dynamic playlist composition. In *Proceedings of SPIE*, volume 6071, pages 24–39.

[Dannenberg and Raphael, 2006] Dannenberg, R. and Raphael, C. (2006). Music score alignment and computer accompaniment. *Communications of the ACM*, 49(8):38–43.

[de Vries et al., 1999] de Vries, A. et al. (1999). *Content and multimedia database management systems*.

[DeCoro et al., 2007] DeCoro, C., Barutcuoglu, Z., and Fiebrink, R. (2007). Bayesian aggregation for hierarchical genre classification. In *Proceedings of the International Conference on Music Information Retrieval*, pages 77–80.

[Deliège et al., 2008] Deliège, F., Chua, B., and Pedersen, T. (2008). High-Level Audio Features: Distributed Extraction and Similarity Search. pages 565–570.

[Dietterich et al., 1997] Dietterich, T., Lathrop, R., and Lozano-Pérez, T. (1997). Solving the multiple instance problem with axis-parallel rectangles. *Artificial Intelligence*, 89(1-2):31–71.

[Dowling and Harwood, 1986] Dowling, W. and Harwood, D. (1986). *Music cognition*. Academic Press New York, NY.

[Eronen and Klapuri, 2000] Eronen, A. and Klapuri, A. (2000). Musical instrument recognition using cepstral coefficients and temporal features. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2.

[Essid et al., 2006] Essid, S., Richard, G., and David, B. (2006). Musical instrument recognition by pairwise classification strategies. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4):1401–1412.

[Ferreira, 2005] Ferreira, P. (2005). A survey on sequence pattern mining algorithms.

[FitzGerald et al., 2006] FitzGerald, D., Cranitch, M., and Coyle, E. (2006). Sound source separation using shifted non-negative tensor factorisation.

[Foote, 1997] Foote, J. (1997). A similarity measure for automatic audio classification. In *Proc. AAAI 1997 Spring Symposium on Intelligent Integration and Use of Text, Image, Video, and Audio Corpora*.

[Foote, 1999] Foote, J. (1999). Visualizing music and audio using self-similarity. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, pages 77–80. ACM.

[Foote and Cooper, 2001] Foote, J. and Cooper, M. (2001). Visualizing musical structure and rhythm via self-similarity. In *Proceedings of the 2001 International Computer Music Conference*, pages 419–422.

[Foote et al., 2002] Foote, J., Cooper, M., and Nam, U. (2002). Audio retrieval by rhythmic similarity. In *Proceedings of the International Conference on Music Information Retrieval*, volume 3, pages 265–266.

[Fujihara et al., 2005] Fujihara, H., Kitahara, T., Goto, M., Komatani, K., Ogata, T., and Okuno, H. (2005). Singer identification based on accompaniment sound reduction and reliable frame selection. In *Proceedings of the International Conference on Music Information Retrieval*, pages 329–336.

[Gelman, 2004] Gelman, A. (2004). *Bayesian data analysis*. CRC press.

[Gillet and Richard, 2007] Gillet, O. and Richard, G. (2007). Supervised and unsupervised sequence modelling for drum transcription. In *Proceedings of the International Conference on Music Information Retrieval*, pages 219–224.

[Grimaldi et al., 2003] Grimaldi, M., Cunningham, P., and Kokaram, A. (2003). A wavelet packet representation of audio signals for music genre classification using different ensemble and feature selection techniques. In *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, pages 102–108. ACM.

[Guo and Siegelmann, 2004] Guo, A. and Siegelmann, H. (2004). Time-warped longest common subsequence algorithm for music retrieval. In *Proceedings of the International Conference on Music Information Retrieval*, pages 258–261.

[Han and Kamber, 2006] Han, J. and Kamber, M. (2006). *Data mining: concepts and techniques*. Morgan Kaufmann.

[Herrera-Boyer et al., 2003] Herrera-Boyer, P., Peeters, G., and Dubnov, S. (2003). Automatic classification of musical instrument sounds. *Journal of New Music Research*, 32(1):3–21.

[Hu et al., 2009] Hu, X., Downie, J., and Ehmann, A. (2009). Lyric text mining in music mood classification. In *Proceedings of the International Conference on Music Information Retrieval*, pages 411–416.

[Hu et al., 2005] Hu, X., Downie, J., West, K., and Ehmann, A. (2005). Mining music reviews: promising preliminary results. In *Proceedings of the International Conference on Music Information Retrieval*, pages 536–539.

[Isaacson, 2005] Isaacson, E. (2005). What you see is what you get: on visualizing music. In *Proceedings of the International Conference on Music Information Retrieval*, pages 389–395.

[Jehan, 2004] Jehan, T. (2004). Perceptual segment clustering for music description and time-axis redundancy cancellation. In *Proceedings of the International Conference on Music Information Retrieval*, pages 124–127.

[Jensen, 2008] Jensen, B. (2008). Exploratory datamining in music.

[Jones and Willett, 1997] Jones, K. and Willett, P. (1997). *Readings in information retrieval*.

[Kameoka et al., 2005] Kameoka, H., Nishimoto, T., and Sagayama, S. (2005). Harmonic-temporal-structured clustering via deterministic annealing EM algorithm for audio feature extraction. In *Proceedings of the International Conference on Music Information Retrieval*, pages 115–122.

[Kim and Whitman, 2002] Kim, Y. and Whitman, B. (2002). Singer identification in popular music recordings using voice coding features. In *Proceedings of the International Conference on Music Information Retrieval*, pages 13–17.

[Knees et al., 2004] Knees, P., Pampalk, E., and Widmer, G. (2004). Artist classification with web-based data. In *Proceedings of the International Conference on Music Information Retrieval*, pages 517–524.

[Knopke, 2004] Knopke, I. (2004). Sound, music and textual associations on the World Wide Web. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 484–488.

[Kolda and Bader, 2009] Kolda, T. and Bader, B. (2009). Tensor decompositions and applications. *SIAM review*, 51(3):455–500.

[Kononenko, 1991] Kononenko, I. (1991). Semi-naive Bayesian classifier. In *Machine Learning-EWSL-91*, pages 206–219.

[Kostek, 2004] Kostek, B. (2004). Musical instrument classification and duet analysis employing music information retrieval techniques. *Proceedings of the IEEE*, 92(4):712–729.

[Kraft et al., 2001] Kraft, R., Lu, Q., and Teng, S. (2001). Method and apparatus for music summarization and creation of audio summaries. US Patent 6,225,546.

[Krumhansl, 2002] Krumhansl, C. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science*, 11(2):45–50.

[Kuo et al., 2005] Kuo, F., Chiang, M., Shan, M., and Lee, S. (2005). Emotion-based music recommendation by association discovery from film music. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 507–510. ACM.

[Lamere, 2008] Lamere, P. (2008). Social tagging and music information retrieval. *Journal of New Music Research*, 37(2):101–114.

[Lampropoulos et al., 2005] Lampropoulos, A., Lampropoulou, P., and Tsihrintzis, G. (2005). Musical genre classification enhanced by improved source separation techniques. In *Proceedings of the International Conference on Music Information Retrieval*, pages 576–581.

[Lenzerini, 2002] Lenzerini, M. (2002). Data integration: A theoretical perspective. In *Proceedings of the 21th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 233–246. ACM.

[Li and Sleep, 2005] Li, M. and Sleep, R. (2005). Genre classification via an lz78-based string kernel. In *Proceedings of the International Conference on Music Information Retrieval*, pages 252–259.

[Li, 2005] Li, T. (2005). A general model for clustering binary data. In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, pages 188–197. ACM.

[Li et al., 2002] Li, T., Li, Q., Zhu, S., and Ogihara, M. (2002). A survey on wavelet applications in data mining. *ACM SIGKDD Explorations Newsletter*, 4(2):49–68.

[Li and Ogihara, 2003] Li, T. and Ogihara, M. (2003). Detecting emotion in music. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 239–240.

[Li and Ogihara, 2004] Li, T. and Ogihara, M. (2004). Music artist style identification by semi-supervised learning from both lyrics and content. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pages 364–367. ACM.

[Li and Ogihara, 2005] Li, T. and Ogihara, M. (2005). Music genre classification with taxonomy. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05)*, pages 197–200.

[Li and Ogihara, 2006] Li, T. and Ogihara, M. (2006). Content-based music similarity search and emotion detection. In *Proceeding of IEEE International Conference on Acoustic, Speech, and Signal Processing*, pages 17–21.

[Li et al., 2003] Li, T., Ogihara, M., and Li, Q. (2003). A comparative study on content-based music genre classification. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pages 282–289. ACM.

[Li et al., 2006] Li, T., Ogihara, M., and Zhu, S. (2006). Integrating features from different sources for music information retrieval. In *Sixth International Conference on Data Mining, 2006. ICDM'06*, pages 372–381.

[Liao et al., 2009] Liao, C., Wang, P., and Zhang, Y. (2009). Mining association patterns between music and video clips in professional MTV. *Advances in Multimedia Modeling*, pages 401–412.

[Liu and Huang, 2002] Liu, C. and Huang, C. (2002). A singer identification technique for content-based classification of MP3 music objects. In *Proceedings of the Eleventh International Conference on Information and Knowledge Management*, pages 438–445. ACM.

[Liu et al., 2003] Liu, D., Lu, L., and Zhang, H. (2003). Automatic mood detection from acoustic music data. In *Proceedings of the International Symposium on Music Information Retrieval*, pages 81–87.

[Liu et al., 2005] Liu, N., Zhang, B., Yan, J., Chen, Z., Liu, W., Bai, F., and Chien, L. (2005). Text representation: From vector to tensor.

[Liu et al., 2008] Liu, Y., Wang, Y., Shenoy, A., Tsai, W., and Cai, L. (2008). Clustering music recordings by their keys. In *Proceedings of the International Conference on Music Information Retrieval*, pages 319–324.

[Logan and Chu, 2000] Logan, B. and Chu, S. (2000). Music summarization using key phrases. In *IEEE International Conf. on Acoustics, Speech, and Signal Processing (ICASSP00)*, volume 2, pages 749–752.

[Logan and Salomon, 2001] Logan, B. and Salomon, A. (2001). A content-based music similarity function. *Cambridge Research Labs-Tech Report*.

[Lübbers, 2005] Lübbers, D. (2005). Sonixplorer: combining visualization and auralization for content-based exploration of music collections. In *Proceedings of the International Conference on Music Information Retrieval*, pages 590–593.

[Lukashevich et al., 2009] Lukashevich, H., Abeßer, J., Dittmar, C., and Grossmann, H. (2009). From multi-labeling to multi-domain-labeling: a novel two-dimensional approach to music genre classification. In *Proceedings of the International Conference on Music Information Retrieval*, pages 459–464.

[Mandel and Ellis, 2008] Mandel, M. and Ellis, D. (2008). Multiple-instance learning for music information retrieval. In *Proceedings of the International Conference on Music Information Retrieval*.

[McKay and Fujinaga, 2004] McKay, C. and Fujinaga, I. (2004). Automatic genre classification using large high-level musical feature sets. In *Proceedings of the International Conference on Music Information Retrieval*, pages 525–530.

[Meng et al., 2005] Meng, A., Ahrendt, P., and Larsen, J. (2005). Improving music genre classification by short time feature integration. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05)*, pages 497–500.

[Meng and Shawe-Taylor, 2005] Meng, A. and Shawe-Taylor, J. (2005). An investigation of feature models for music genre classification using the support vector classifier. In *Proceedings of the International Conference on Music Information Retrieval*, pages 604–609.

[Nam and Berger, 2001] Nam, U. and Berger, J. (2001). Addressing the "same but different-different but similar" problem in automatic music classification. In *Proceedings of the International Conference on Music Information Retrieval*, pages 21–22.

[Ni et al., 2008] Ni, K., Paisley, J., Carin, L., and Dunson, D. (2008). Multi-task learning for analyzing and sorting large databases of sequential data. *IEEE Transactions on Signal Processing*, 56(8):3918–3931.

[Norowi et al., 2005] Norowi, N., Doraisamy, S., and Wirza, R. (2005). Factors affecting automatic genre classification: an investigation incorporating non-western musical forms. In *Proceedings of the International Conference on Music Information Retrieval*, pages 13–20.

[Orio, 2006] Orio, N. (2006). *Music retrieval: A tutorial and review*.

[Pampalk et al., 2005] Pampalk, E., Flexer, A., and Widmer, G. (2005). Improvements of audio-based music similarity and genre classification. In *Proceedings of the International Conference on Music Information Retrieval*, pages 628–633.

[Pampalk et al., 2002] Pampalk, E., Rauber, A., and Merkl, D. (2002). Content-based organization and visualization of music archives. In *Proceedings of the tenth ACM international conference on Multimedia*, pages 570–579. ACM.

[Peeters, 2007] Peeters, G. (2007). Sequence representation of music structure using higher-order similarity matrix and maximum-likelihood approach. pages 35–40.

[Peeters et al., 2002] Peeters, G., La Burthe, A., and Rodet, X. (2002). Toward automatic music audio summary generation from signal analysis. In *Proc. International Conference on Music Information Retrieval*, pages 94–100.

[Peng et al., 2007] Peng, W., Li, T., and Ogihara, M. (2007). Music clustering with constraints. In *Proceedings of the International Conference on Music Information Retrieval*, pages 27–32.

[Pienimäki and Lemström, 2004] Pienimäki, A. and Lemström, K. (2004). Clustering symbolic music using paradigmatic and surface level analyses. In *Proceedings of the International Conference on Music Information Retrieval*, pages 262–265.

[Pryakhin, 2006] Pryakhin, A. (2006). Similarity search and data mining techniques for advanced database systems.

[Rabiner and Juang, 1993] Rabiner, L. and Juang, B. (1993). *Fundamentals of speech recognition*.

[Rauber et al., 2002a] Rauber, A., Pampalk, E., and Merkl, D. (2002a). Content-based music indexing and organization. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 409–410. ACM.

[Rauber et al., 2002b] Rauber, A., Pampalk, E., and Merkl, D. (2002b). Using psycho-acoustic models and self-organizing maps to create a hierarchical structuring of music by sound similarity. In *Proceedings of the International Conference on Music Information Retrieval*, pages 71–80.

[Reed and Lee, 2006] Reed, J. and Lee, C. (2006). A study on music genre classification based on universal acoustic models. pages 89–94.

[Rodriguez et al., 2008] Rodriguez, A., Dunson, D., and Gelfand, A. (2008). The nested Dirichlet process (with discussion). *Journal of American Statistical Association*, 103:1131–1144.

[Sagayama et al., 2004] Sagayama, S., Takahashi, K., Kameoka, H., and Nishimoto, T. (2004). Specmurt anasylis: A piano-roll-visualization of polyphonic music signal by deconvolution of log-frequency spectrum. In *ISCA Tutorial and Research Workshop (ITRW) on Statistical and Perceptual Audio Processing*.

[Sandvold et al., 2004] Sandvold, V., Gouyon, F., and Herrera, P. (2004). Percussion classification in polyphonic audio recordings using localized sound models. In *Proceedings of the International Conference on Music Information Retrieval*, pages 537–540.

[Schnitzer et al., 2009] Schnitzer, D., Flexer, A., Widmer, G., and Linz, A. (2009). A filter-and-refine indexing method for fast similarity search in millions of music tracks. pages 537–542.

[Scholz and Ramalho, 2008] Scholz, R. and Ramalho, G. (2008). Cochonut: recognizing complex chords from mIDI guitar sequences. In *Proceedings of the International Conference on Music Information Retrieval*, pages 27–32.

[Shao et al., 2004] Shao, X., Xu, C., Wang, Y., and Kankanhalli, M. (2004). Automatic music summarization in compressed domain. In *IEEE International Conf. on Acoustics, Speech, and Signal Processing (ICASSP04)*, pages 261–264.

[Shen et al., 2006a] Shen, J., Cui, B., Shepherd, J., and Tan, K. (2006a). Towards efficient automated singer identification in large music databases. In *Proceedings of the 29th Annual International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 59–66. ACM.

[Shen et al., 2006b] Shen, J., Shepherd, J., and Ngu, A. (2006b). InMAF: indexing music databases via multiple acoustic features. In *Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data*, pages 778–780. ACM.

[Slaney et al., 2008] Slaney, M., Weinberger, K., and White, W. (2008). Learning a metric for music similarity. In *Proceedings of the International Conference on Music Information Retrieval*, pages 313–381.

[Sun et al., 2006] Sun, J., Tao, D., and Faloutsos, C. (2006). Beyond streams and graphs: dynamic tensor analysis. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 374–383. ACM.

[Tan et al., 2006] Tan, P., Steinbach, M., and Kumar, V. (2006). *Introduction to data mining*. Pearson Addison Wesley Boston.

[Teh et al., 2006] Teh, Y., Jordan, M., Beal, M., and Blei, D. (2006). Hierarchical dirichlet processes. *Journal of American Statistical Association*, 101(476):1566–1581.

[Thrun and O'Sullivan, 1996] Thrun, S. and O'Sullivan, J. (1996). Discovering structure in multiple learning tasks: The TC algorithm. In *International Conference on Machine Learning*, pages 489–497.

[Torrens et al., 2004] Torrens, M., Hertzog, P., and Arcos, J. (2004). Visualizing and exploring personal music libraries. In *Proceedings of the International Conference on Music Information Retrieval*, pages 421–424.

[Tsai et al., 2004] Tsai, W., Rodgers, D., and Wang, H. (2004). Blind clustering of popular music recordings based on singer voice characteristics. *Computer Music Journal*, 28(3):68–79.

[Tsai et al., 2003] Tsai, W., Wang, H., and Rodgers, D. (2003). Automatic singer identification of popular music recordings via estimation and modeling of solo vocal signal. In *Eighth European Conference on Speech Communication and Technology*, pages 2993–2996.

[Tsoumakas et al., 2008] Tsoumakas, K., Kalliris, G., and Vlahavas, I. (2008). Multi-label classification of music into emotions. In *Proceedings of the International Conference on Music Information Retrieval*, pages 325–330.

[Typke et al., 2005] Typke, R., Wiering, F., and Veltkamp, R. (2005). A survey of music information retrieval systems. In *Proceedings of the International Conference on Music Information Retrieval*, pages 153–160.

[Tzanetakis and Cook, 2001] Tzanetakis, G. and Cook, P. (2001). Marsyas3D: a prototype audio browser-editor using a large scale immersive visual and audio display. In *Proceeding of International Conference on Auditoy Display (ICAD)*, pages 250–254.

[Tzanetakis and Cook, 2002] Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293–302.

[Wang et al., 2009] Wang, F., Wang, X., Shao, B., Li, T., and Ogihara, M. (2009). Tag integrated multi-Label music style classification with hypergraph. In *Proceedings of the International Conference on Music Information Retrieval*, pages 363–368.

[Wold et al., 1996] Wold, E., Blum, T., Keislar, D., and Wheaten, J. (1996). Content-based classification, search, and retrieval of audio. *IEEE multimedia*, 3(3):27–36.

[Xiao et al., 2008] Xiao, L., Tian, A., Li, W., and Zhou, J. (2008). Using a statistic model to capture the association between timbre and perceived tempo. In *Proceedings of the International Conference on Music Information Retrieval*, pages 659–662.

[Xu et al., 2005] Xu, C., Maddage, N., and Shao, X. (2005). Automatic music classification and summarization. *IEEE Transactions on Speech and Audio Processing*, 13(3):441–450.

[Xu et al., 2004] Xu, C., Shao, X., Maddage, N., Kankanhalli, M., and Tian, Q. (2004). Automatically summarize musical audio using adaptive clustering. In *2004 IEEE International Conference on Multimedia and Expo, 2004. ICME'04*, pages 2063–2066.

[Xue et al., 2007] Xue, Y., Liao, X., Carin, L., and Krishnapuram, B. (2007). Multi-task learning for classification with Dirichlet process priors. *The Journal of Machine Learning Research*, 8:35–63.

[Yang et al., 2006] Yang, Y., Liu, C., and Chen, H. (2006). Music emotion classification: a fuzzy approach. In *Proceedings of the 14th annual ACM international conference on Multimedia*, pages 81–84. ACM.

[You and Dannenberg, 2007] You, W. and Dannenberg, R. (2007). Polyphonic music note onset detection using semi-supervised learning. pages 279–282.

[Zhang, 2003] Zhang, T. (2003). Automatic singer identification. In *Proceedings of the International Conference on Multimedia and Expo*, pages 33–36.

[Zhang et al., 1998] Zhang, T., Kuo, C., et al. (1998). Hierarchical system for content-based audio classification and retrieval. In *Conference on Multimedia Storage and Archiving Systems III, SPIE*, volume 3527, pages 398–409.