

面向推荐系统的音乐特征抽取

张 燕^{1,2}, 唐振民¹, 李燕萍³

ZHANG Yan^{1,2}, TANG Zhenmin¹, LI Yanping³

1. 南京理工大学 计算机科学与技术学院, 南京 210094

2. 金陵科技学院 信息技术学院, 南京 211169

3. 南京邮电大学 通信工程学院, 南京 210046

1. College of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China

2. College of Information Technology, Jinling Institute of Technology, Nanjing 211169, China

3. College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210046, China

ZHANG Yan, TANG Zhenmin, LI Yanping. Music feature extraction method for recommendation system. Computer Engineering and Applications, 2011, 47(5): 130-133.

Abstract: Music recommendation system is to provide needful music for users from the candidate music library based on analyzing the user's browsing history data. The classification of whole database according to music style is the key point of this system. Then this paper proposes a novel music feature for achieving music classification. Firstly the regular data sets of music feature are constructed. Then the attribute of data sets is reduced based on fractal dimension theory and the recommendation features are obtained for each music. Also a new distance measure is given for the similarity of this recommendation feature. During the experiment based on database including six styles of music, simulation results demonstrate the effectiveness of the proposed music recommendation feature and the new distance. Compared with the study of current content-based music retrieval, music recommendation feature can greatly reduce the requirement of database memory and provide a promising application for the network exploitation of music recommendation system.

Key words: music recommendation system; fractal dimension; feature extraction; vector similarity

摘 要: 音乐推荐系统是指根据用户的历史浏览数据, 从候选库中推荐给用户可能喜欢的音乐的一种新型网络服务。该系统的关键在于需要对整个数据库按照音乐风格进行分类, 基于此提出一种新的音乐特征处理方法来完成音乐库分类, 以有效实现音乐推荐。该方法首先为候选音乐库构建常规的音乐特征数据集, 然后基于分形理论对数据集进行属性约简, 获取每一首音乐的推荐特征向量, 并且依据特征向量的特点, 定义了一种新的距离度量方法。在包含六种风格的音乐数据库的实验中, 仿真结果证明了提出的音乐推荐特征和距离度量的有效性, 与现有的基于内容的音乐检索研究相比, 音乐推荐特征的使用极大地降低了对数据库存储量的需求, 对音乐推荐系统的网络开发具有很好的应用价值。

关键词: 音乐推荐系统; 分形维; 特征抽取; 向量相似度

DOI: 10.3778/j.issn.1002-8331.2011.05.039 文章编号: 1002-8331(2011)05-0130-04 文献标识码: A 中图分类号: TP391.42

基于用户的网页查阅历史及网页本身的内容, 获取用户的个性化需求, 并依此推荐给相关信息, 实现为主动导航, 这种推荐系统使得用户从Internet海量信息中获取所需知识更为方便。

基于内容的音乐推荐系统是音乐或娱乐网站大力发展的新主题, 但是与文本内容不同的是音乐具有更大的数据量和难以描述的推荐特征; 前者将导致算法的时间复杂度难以达到在线处理的要求, 后者使搜索具有很大的模糊性: 用户喜欢某首歌可能是因为歌手、歌词、节奏、旋律等多种因素, 甚至自己也不能清楚描述其原因所在, 这使推荐系统的音乐特征难以确定。

音乐推荐系统中基于内容的音乐特征(以下简称音乐推荐特征)并不完全等同于音频检索、分割中的音乐特征, 后者需要对某一音乐信息的完整表达, 而音乐推荐系统需要获取其整体性特征为个性化推荐提供依据。因此, 音乐推荐特征实际可以看作个体音乐相对于整个音乐库的个性描述, 此外还要求特征维数不能太大, 以保证在线处理的能力。

介绍的算法首先通过设定音乐的内容特征, 构建音乐库的特征记录集; 之后采用分形理论对记录集进行属性约简, 得到 n 维矢量构成音乐推荐特征向量, 最后定义向量之间的距离度量获得不同音乐之间的相似度, 实现音频的归类和推荐。

基金项目: 江苏省现代教育技术研究所课题(the Issue of Jiangsu Institute of Modern Educational Technology No.2007-R-4704)。

作者简介: 张燕(1969—), 女, 博士研究生, 副教授, 主要研究领域为基于内容的音频特征识别与检索技术, 计算机教育; 李燕萍(1983—), 女, 博士, 讲师, 主要研究领域为语音处理, 语音转换。E-mail: zy@jit.edu.cn

收稿日期: 2010-08-26; **修回日期:** 2010-11-17

1 基于内容的音乐特征

有很多特征被提取出来用于描述音乐,包括频域能量、子带能量、频域中心、基音频率、MFCC 系数(Mel-Frequency Cepstral Coefficient)等^[1],其主要思想是对音乐音高(频率)、音长(节奏)、响度(重音)的刻画等。这些特征需要将音频分割成等时间的帧、加窗,有时为了保证帧间特征的平滑还需要设置帧交叠,经过傅里叶变换得到帧频谱 $F(w)$ 后,可以计算以下音乐特征:

$$\text{频域能量: } E = \int_0^w |F(w)|^2 dw, \bar{w} = fs/2。$$

子带能量: $SE = \int_{w \in sb_i} |F(w)|^2 dw$, 将频谱分为若干个子带 sb_i , 以统计频域能量的分布特征。

频率中心: $FC = \int_0^w w |F(w)|^2 dw / E$, 该特征是度量音频亮度的指标。

带宽: $BW = \sqrt{\int_0^w (w - FC)^2 |F(w)|^2 dw / E}$, 一般用于衡量音乐的频域范围。

基音频率: $F0$, 第一个谐波频率, 用于衡量音调高低。

MFCC 系数: 主要表达声道或乐器的音色特征。

基于上述特征还可以推导出音乐帧的其他特征, 如静音比例、子带能量均值、带宽均值、基音频率标准方差、谐成份比例、相邻帧 MFCC 差值等。之后还提出了一些新音乐特征描述方法, 如以声谱图来概略地描述一段音乐^[2], 该方法选取每一帧中最大的 K 个频率值构建“特征识别矩阵”, 依据矩阵的相似性进行音乐检索。此外还有使用模糊粗糙集模型的音乐数据约简算法^[3], 以及 MDCT 频谱熵压缩域音频指纹算法等^[4], 这些方法都在音乐检索中得到了验证。

2 音乐记录集的分维计算

基于内容的检索希望依据某一曲或一些音乐, 得到相同或相近的其他作品, 这是一个分类和查询的过程。而推荐系统是从现有音乐库中检索出相似的其他作品, 这种相似特征带有相对性, 音乐的相似不是指在某个时间段具有相同的音乐特征, 而是总体感觉接近。因此, 从特征选择角度来看, 音乐推荐比基于内容的检索更强调调整体的概念, 这也符合实际应用需求。

基于以上分析, 构建音乐库特征数据集如下: 对所有音乐进行等长分帧。提取每一首音乐的音乐特征(如第1章所述)构成一条记录, 所有音乐特征构成了一个音乐记录集。对一个数据集而言, 并非所有属性都是记录分类所必须的, 分类无关或弱相关的属性为冗余属性。为了删除原始音乐记录集的冗余属性, 需要利用分形维数进行属性约简。

分形来自于对自然、类自然的描述, 在音乐处理中也得到了应用, 如 Bigerelle M 等^[5]利用音乐的分形维数作为音乐特征对多种音乐进行分类, 李坚等^[6]提出利用分形几何抽取音频特征的全局化音频检索, 在学习阶段计算音频数据库中每个音频的分维作为特征向量。但这些研究是将分形应用于具体一首音乐或一首音乐的局部特征, 而分形的本质是总体与部分的相似, 因此音乐特征抽取的记录集也具有分形特征。

一个数据集是否具有分形特征尚未有理论上的鉴定, 但是通过对具体数据集的分形维数计算可以得到反演认定, 这

也是利用分形理论对记录集进行属性约简的基础, 研究证明, 分形理论应用到拥有较多属性的数据集上可以达到更好的属性约简效果^[7-10]。

实际求解一个数据集的分形维数常常采用计算记录集的盒维来近似得到, 其具体过程如下: 设数据集 $s = \{A, E, d\}$, 其中 A 表示拥有 p 个属性的属性集 $\{A_1, A_2, \dots, A_p\}$, d 表示分类属性, E 表示包含 n 个元组的对象集, 依据 Traina 等人^[7]的研究, 如果将 E 映射到 p 维空间, 并将每一维 $1/r$ 等分 ($r=1/2, 1/4, 1/8, \dots$) 后得到 $(1/r)^p$ 个单元格, 依属性集顺序从 0 开始编号每一个 p 维单元格, 则第 i 条记录 $Ri = \{Ri1, Ri2, \dots, Rip\}$ 在 p 维空间中落入的单元格顺序编号可以基于 $(Ri1/R1, Ri2/R2, \dots, Rip/Rp)$ 得到, 其中 $R1, R2, \dots, Rp$ 分别为数据集的属性 A_1, A_2, \dots, A_p 的值域范围被 $1/r$ 等分之后的数值, 记第 i 个单元格落入的点数为 $C(r, i)$, 令 $S(r) = \sum_i C(r, i)^2$, 则数据集的分形维数可以定义为:

$$D(s) = -\frac{\partial \log(S(r))}{\partial \log r} \quad (1)$$

对于自相似的数据集, 式(1)是常数, 而对于实际数据集, 在以 $\log r$ 为横坐标, 以 $\log(S(r))$ 为纵坐标的两维坐标系中描点, 其曲线的近似斜率即为数据集的分形维。研究证明数据集的分形维数是数据集的固有维数的一个精确度量^[7-9]。由式(1)可知, 分形维数的计算并不需要分类属性的参与, 而分类属性常常应用在检验过程中。

3 构建音乐推荐特征

设音乐记录集的原有属性为 m , 通过计算出音乐记录集的盒维 n , 可以认定原记录集的固有维数为 n , 即数据集有 $(m - n)$ 个冗余属性, 与传统的维数约简过程中采用剔除冗余属性或选择候选属性的方法提取数据集的固有维数不同, 提出基于每条记录进行独立处理的策略。首先, 音乐记录集中每一属性并不同于传统数据集中的严格定义, 根据一首音乐的等时间段划分获取的某一特征仅仅反映的是该音乐在一个时段的特点。另一方面, 两首相似的音乐是因为具有更多的相似特征, 并且相似特征强度越大, 其相似性表现得愈明显, 文献[2]就以声波强度进行音乐特征的构建。因此, 对每一首音乐的原有 m 个特征选择数值最大的前 n 个特征, 并依据时间顺序构成推荐特征。这一方法同时也提高了特征集的抽取效率。

4 特征向量之间的度量

两个音乐推荐特征向量之间的相似性并不等同于传统向量之间的度量, 因为传统向量的每一个分量表达相似性的能力相当, 因此采用夹角余弦^[11]可以反映出两个传统向量之间的相似度, 对于两个特征向量 $X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_n\}$, 其夹角余弦的定义为:

$$\rho = \frac{\text{cov}(x, y)}{\rho_x \rho_y} = \frac{\sum_{i=1}^n (x_i y_i)}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (2)$$

夹角余弦的取值在 $[0, 1]$ 范围, 当其值为 1 时, 两个向量是完全一样的。

但就音乐推荐特征向量而言, 两个向量中序号相同的一对分量并不表示相同的帧位置, 但向量的顺序号反映了同一

首音乐中帧的顺序,其数值的接近与否代表了音乐的相似程度。因此式(2)仅从一定程度上对音乐推荐特征向量之间的相似性进行了描述。

音乐的相似性也可能在不同的时间顺序上有相同或相似的旋律等特点,反映在特征向量上可能表现为两个向量有很多相同的分量值,但它们并不是彼此对应的,如以下向量: $x=\{1,1,1,5,5,5\}$, $y=\{5,5,5,1,1,1\}$,可以看出这两个音乐推荐特征向量有较大的相似性,但由式(2)算得的相似结果仅为0.385,甚至没有特征 $x=\{1,1,1,1,1,1\}$, $y=\{2,2,2,2,2,2\}$ 得到的结果大(0.4)。因此比较两个音乐推荐特征向量之间的相似性,需要忽略它们的顺序,当二者有更多相同或接近数值的分量时,定义其为相似。针对这一特性定义两个音乐推荐特征向量的距离 d_2 为一个向量间所有分量与另一向量中所有分量最接近值之差的绝对值之和。 d_2 的获得需要遍历向量的所有特征值,两向量距离的伪代码如下。

```
Input: 向量  $X=\{x_1, x_2, \cdots, x_n\}$ ,  $Y=\{y_1, y_2, \cdots, y_n\}$ ;
Output: 距离  $d_2$ 。
Process:
  初始化:  $d_2=0$ ;
  循环: for (i=1 to n) { //n 为向量维数
     $d_i=\text{abs}(x_i - y_i)$ ; //基准距离
    flag=i; //最小距离标记
    循环: for (j=1 to n-i+2) {
       $d_j=\text{abs}(x_i - y_j)$ ;
      if  $d_j=0$  {
        flag=j;
        exit;
      }
      if  $d_j < d_i$  { //更新更小的距离
         $d_i=d_j$ ;
        flag=j;
      }
    }
    if flag < n - i + 2
       $y_{\text{flag}}=y_{n-i+2}$ ;
     $d_2=d_2+d_i$ ; //总距离累加
  }
```

由伪代码可以看出,计算距离的时间复杂度为 $O(n^2)$, n 代表向量的维数,而实际 n 值一般在几百以内,因此其耗时是很小的。对由伪代码获得的距离进行数据归一化处理,最终定义为: $d_2=1-e^{-d_2}$,基于式(2)定义另一距离为: $d_1=1-\rho$ 。由此得到两个音乐推荐特征向量之间的距离为:

$$d=\frac{d_1+d_2}{2}$$

(3)

基于式(3)得到上文中提到的第一对向量之间的距离为0.307 5,而第二对向量之间的距离为0.797 5,符合本文预期。

5 仿真实验

5.1 实验素材及方法

为了检验提出的推荐特征的有效性,分别选取wav格式的音乐683首,包括男女生流行音乐独唱歌曲、女声小合唱、钢琴演奏、古筝独奏、二胡独奏等六种风格的音乐,数据采用双声道、采样率为22.05 kHz、存储精度为16位格式。所选音乐要求时间比较接近,全部在3.6 min至5.2 min之间。

实验主要抽取了以下三个特征进行分析:频域能量,基音频率标准方差以及基于声谱图的识别矩阵的第一行^[2],即每一帧中强度最大的一个频率值。三个特征分别代表了音乐的原始特征、初步统计特征、较高度度的抽象特征。

为了减少音乐库记录集的初始属性数,实验中对每首音乐划分为6个580 ms的音乐片段,相邻片段之间间隔相同的时间,并保证其均匀地分布在整首音乐上,再对每段加23 ms的Hamming窗形成帧,相邻帧有50%重叠,这样一首音乐可以提取300个特征。

为了防止式(3)运算过程中,数据本身的较大差异使距离权重偏大或偏小,对所选特征值都进行了归一化处理。

5.2 实验结果

首先利用式(1)计算数据集的分形维数,图1是基于频域能量形成的初始数据集K1在迭代200次下的盒维走势图(因数据差别较大,纵横坐标比例不同,且横轴是单元格个数 r ,而不是 $\log(r)$),从盒维比较光滑的走势及总体斜率相似的特点,可以断定K1数据集具有良好的分形特征。对K1计算得到分形维数为138。同样的计算,基音频率标准方差的初始数据集K2的分形维数为122,每帧强度最大频率值的初始数据集K3的分形维数为175。

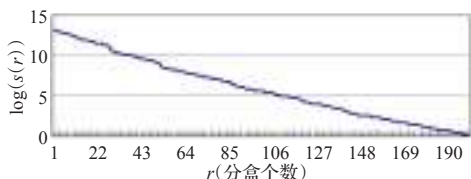


图1 200次迭代下K1集盒维走势图

分别从K1的每一条记录选取TOP138个值构成基于频域能量的特征数据集F1,F1是拥有138个属性683条记录的数据集,同样的F2(122X683)、F3(175X683),利用k-mean聚类方法,采用式(3)的距离定义,分别对记录集F1,F2,F3进行6分类聚类,得到的结果见表1。

表1 基于推荐特征集的分类结果表

	女独	男独	女合	钢琴	古筝	二胡
	A1	A2	A3	A4	A5	A6
原数量	335	200	32	60	28	28
F1 聚类	323	200	26	57	20	25
	6.43	3.41	9.41	5.45	3.44	3.45
F2 聚类	318	198	25	60	20	23
	7.43	5.41	12.41	5.45	3.46	3.45
F3 聚类	321	200	27	58	22	23
	5.43	2.41	12.41	3.45	2.46	3.45
K1 数据	326	200	29	60	22	25
	3.43		9.41	3.45	3.46	3.45

表1数据中形如“6.43”代表将6首A3类(女声小合唱)错分为本类。从表1可以看到本文提出的推荐特征是非常有效的,对原始特征、初步统计特征、较高度度的抽象特征都有良好的效果,说明提出的新方法得到了很好的效果。

为了验证约简后的数据集是否是必须的分类数据集,即没有或仅有很少的冗余属性,对K1数据集分别抽取了每一条记录的TOP30, TOP50, TOP70, TOP90, TOP110, TOP130,

TOP150个值构成SK1~SK7个数据集,进行的 k -mean聚类结果见表2。

表2 基于频域能量的特征在不同属性数下的分类结果表

	女独	男独	女合	钢琴	古筝	二胡
	A1	A2	A3	A4	A5	A6
SK1 聚类	113 73Q	38 82Q	15 67Q	30 75Q	13 79Q	15 83Q
SK2 聚类	128 69Q	55 59Q	18 59Q	33 80Q	15 65Q	18 84Q
SK3 聚类	153 44Q	93 50Q	18 52Q	38 77Q	16 58Q	19 65Q
SK4 聚类	216 31Q	105 38Q	22 50Q	46 59Q	19 45Q	19 33Q
SK5 聚类	283 30Q	187 15Q	26 16Q	50 29Q	21 12Q	21 13Q
SK6 聚类	316 6.43	200 4.41	26 15.41	55 5.45	20 5.44	22 5.45
SK7 聚类	326 6.43	200	26 9.41	57 5.45	20 3.44	25 3.45
					3.46	

表2中“Q”代表被错分音乐类别,分析表1可以看出约简后的数据集中的属性并没有太多的冗余。依据表2和表1最后一行数据所绘制的折线图(图2)可以直观地发现这一特点。

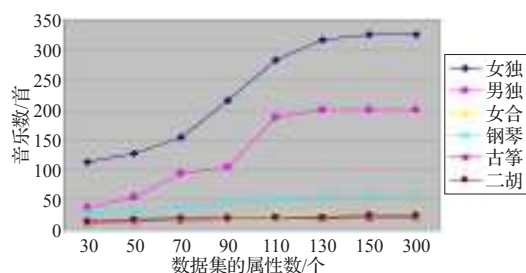


图2 不同属性数下音乐特征集正确分类结果图

6 结语

提出了对整个音乐集进行统一处理的研究思路,这事实上将无穷搜索变成了在有限范围的查找。相比于基于内容的

音乐检索等研究中需要用特征矩阵描述一首音乐,推荐特征向量的数据量仅是它们的几分之一,这对网络服务的实时性是至关重要的。但实验仅仅查询出同类音乐,能否结合多个音乐特征更为精确地推荐出相似的音乐还值得进一步研究。

在实际一个音乐或娱乐网站中的音乐数据库需要不停地更新,因此进行在线的音乐推荐特征抽取是不可能的,一种方案是当整个音乐库中的音乐更新到某种程度后,系统进行一次离线推荐特征抽取并更新每一首音乐的推荐特征,这就会影响最新音乐的及时推荐。如何保证在线服务的前提下有效地实现增量更新,也是下一步将要关注的工作。

参考文献:

- [1] 卢坚,陈毅松,孙正兴,等.语音/音乐自动分类中的特征分析[J].计算机辅助设计与图形学学报,2002,14(3):233-237.
- [2] 孔旭,关佑.以声谱图相似度为度量的波形音乐检索[J].计算机工程与应用,2009,45(13):136-141.
- [3] 李晓丽,杜振龙.模糊粗糙集在音频检索中的应用[J].计算机工程与应用,2010,46(15):124-126.
- [4] 刘亚多,李伟,李晓强,等.压缩域鲁棒音乐指纹算法研究[J].电子学报,2010(5):1172-1176.
- [5] Bigerelle M, Iost A. Fractal dimension and classification of music[J]. Chaos, Solitons and Fractals, 2000, 11(14):2179-2192.
- [6] 李坚,毛先领,文贵华.基于分形特征的音频检索[J].计算机工程,2008,34(6):211-213.
- [7] Traina C, Jr, Traina A. Fast features selection using fractal dimension[C]//Proc of XV Brazilian Database Symposium on Database. Berlin: Springer, 2000:158-171.
- [8] 鲍玉斌,王琢,孙怀良,等.一种基于分形维的快速属性选择算法[J].东北大学学报,2003,24(6):527-530.
- [9] 郭平,陈其鑫,王燕霞.基于分形维数的属性约减[J].计算机科学,2007,34(9):189-190.
- [10] 闫光辉,李怀战.两阶段无监督顺序前向分形属性规约算法[J].计算机研究与发展,2008,45(11):1955-1964.
- [11] Godin R, Missaoui R, Alaoui H. Incremental concept formation algorithms based on Galois (concept) lattices[J]. Computational Intelligence, 1995, 11(2):246-267.

(上接91页)

参考文献:

- [1] Looney C G. Fuzzy Petri nets and application[C]//Tzafestas S G, Venetsanopoulos A N. Fuzzy Reasoning in Information, Decision and Control Systems. Norwell, MA: Kluwer Academic Publishers, 1994:511-527.
- [2] 鲍培明.模糊Petri网模型的反向推理算法[J].南京师范大学学报:工程技术版,2003,3(3):21-25.
- [3] 汪洋,林闯,曲扬,等.含有否定命题逻辑推理的一致性模糊Petri网模型[J].电子学报,2006(11):1955-1960.
- [4] 黄光球,任大勇.基于双枝模糊决策与模糊Petri网的攻击模型[J].计算机应用,2007,27(11):2689-2693.
- [5] 史开泉.双枝模糊集(I)[J].山东工业大学学报,1998,28(2):

127-134.

- [6] 史开泉,李岐强.双枝模糊决策与决策识别问题[J].中国工程,2001,3(1):71-77.
- [7] 蒋仁言,左明健.可靠性模型与应用[M].北京:机械工业出版社,1992:1-2.
- [8] 高梅梅,吴智铭.模糊时间Petri网的时间推理及其在过程监测中的应用[J].控制与决策,2001,16(3):371-373.
- [9] 王圣金.基于蒙特卡洛仿真的液压系统动态可靠性研究[D].南京:东南大学,2006.
- [10] 黄光球,王金成.具有动态可靠性的模糊时间Petri网攻击模型[J].计算机工程与应用,2010,46(18):110-115.
- [11] 黄光球,王金成.基于双枝模糊集的一致性模糊变权Petri网攻击模型[J].计算机应用,2009,29(2):529-535.