音乐情感的自动识别

蒋旻隽, 周昌乐*, 黄志刚
(厦门大学 信息科学与技术学院, 福建省仿脑智能系统重点实验室, 福建 厦门 361005)

摘要: 音乐与情感有着非常密切的联系, 发展针对音乐的情感识别系统, 对于计算机音乐的研究与发展有着深远的意义. 提出了一种基于 PAD (Pleasure arousal dominance) 模型以及基因表达式编程(GEP) 算法的音乐情感自动识别方法. 在众多音乐特征元素中抽取与情感关系密切的 6 个特征, 并且采用 PAD 模型来描述音乐中的情感, 在此基础上使用 GEP 算法实现对简单乐曲中单一情感的自动识别. 从实验结果分析, 本系统能够达到一个比较理想的识别效果和较低的识别误差.

关键词: 情感识别; PAD 模型; 音乐特征; GEP 算法
中图分类号: TP 391 文献标识码: A 文章编号: 0438-0479(2010)06-0798-05

音乐是情感的语言, 音乐与情感有着非常密切的联系[1]. 因此, 音乐情感对音乐的理解、欣赏、创作等都有着重要的影响. 多媒体与人工智能技术的结合已经成为目前的一个研究热点, 计算机音乐就是在这样的背景下应运而生的. 对于计算机音乐的研究来说, 发展针对音乐的自动情感识别系统, 对于音乐检索、音乐自动理解、自动作曲等研究工作都有着重要的影响和意义.

然而, 自动识别音乐中的情感存在着诸多的困难. 情感是非常主观的东西, 在不同的人之间存在较大的差异, 并且很难将其量化. 对于音乐情感识别、检测和分类的研究, 较为常用的一种方法是首先将音乐情感划分为有限的几类, 然后用特征向量来描述乐曲, 最后使用分类方法进行情感分类. 使用的分类方法主要有支持向量机(SVM)[2], 高斯混合模型(GMM)[3], 最近邻(KNN) 算法[4] 等等. 对于音乐情感的表达, 比较常用的描述方法有 Thayer 的二维模型, 以及 Hevner 的音乐情感环. 文献[5] 基于 Hevner 情感环, 提出了一个 8 维的语义标签特征向量来表示音乐中表现出来的情感, 并且使用了一个新的方法来衡量不同音乐之间的情感相似度; 又如文献[6] 根据模糊逻辑原理和 Hevner 情感环提出了一个基于形容词的音乐情感描述环来进行情感计算.

本文采用的情感描述方法与以往的方法不同. 我们不是简单地将音乐划分成几个类别, 也不是基于常用的 Hevner 情感环或者 Thayer 的二维模型, 而是采用 PAD(Pleasure arousal dominance) 模型, 对于每首乐曲, 用一个 PAD 值来描述其中的情感; 经过分析, 我们从众多音乐特征元素中找出与情感有关 6 个特征值, 它们组成了乐曲的特征向量; 最后在此基础上, 使用基因表达式编程(GEP) 算法, 通过对训练样本的学习, 最终得到 PAD 3 个分量的公式. 对于任意一首新的乐曲, 只需要抽取出相应的特征向量, 就可以根据这 3 个公式得到属于该乐曲的 PAD 值, 由此来描述其情感状态. 实验表明, 测试样本的误差值能够控制在一个合理的范围内.

1 基于 PAD 的音乐情感描述

为了方便计算机处理, 主观的音乐情感必须事先进行量化. 为了实现对情感的准确测量, 在心理学领域, 人们已经提出了一些情感测量理论和具体的测量方法. 通过对比分析, 本文中我们采用 PAD 模型[7].

PAD 三维情感模型是由 Mehrabian 和 Russell 于 1974 年提出的维度观测量模型. 该模型将情感分为愉悦度、激活度和优势度 3 个维度. 其中, P 代表愉悦度, 表示个体情感状态的正负特性; A 代表激活度, 表示个体的神经生理激活水平; D 代表优势度, 表示个体对情景和他人的控制状态. 通过这 3 个维度的值可以表示各种具体的情感.

研究表明, 利用 P、A、D 3 个维度可有效地解释人类的情感. 例如, Mehrabian 等利用这 3 个维度解释了

收稿日期: 2010-03-19
基金项目: 国家自然科学基金资助项目(60975076)
* 通讯作者: dozero@xmu.edu.cn

其他 42 种情感量表中的绝大部分变异. 而且这 3 个维度并不限于描述情感的主观体验, 它与情感的外部表现、生理唤醒具有较好的映射关系.

与其他的情感表示方法相比, PAD 有其特点与优势. 首先对于 Thayer 模型来说, 只有两个维度, 在表述情感的丰富性上有所欠缺; 而 Hevner 情感环, 对于情感的分类以及描述非常详细, 除 8 维的情感环以外, 每一个类别又细分为几个子类, 例如文献[8] 中建立的基于 Hevner 情感环的语义模型, 但考虑到这种分类过于细致, 而针对只具有单一情感的简单乐曲无需过于复杂的语义模型, 本文中, 我们将使用 P、A、D 这 3 个值来描述音乐的情感状态. 值得注意的是, 我们选用的音乐样本都是目前的流行音乐, 所以情感相对于大型的音乐作品要更为简单, 不会出现复杂的复合情感, 基本上都认为只有一个单一的情感包含在乐曲当中. 我们做以下定义:

定义 **1**　对于一首音乐的情感, 它可以由 PAD 的 3 个分量来描述, 即 $\mathrm{Emoyion}(M_i) = \{P_i, A_i, D_i\}$, 其中 $P_i \in [-1, 1]$, $A_i \in [-1, 1]$, $D_i \in [-1, 1]$.

2　音乐特征的抽取

收集音乐样本是我们工作的前提, 本文将采用 MIDI 格式的当代流行音乐作为样本. 收集到的样本不能直接用于程序当中, 必须经过一系列预处理, 主要包括两方面: MIDI 主旋律音轨的提取; 音乐特征的确定与抽取.

2.1　MIDI 音乐主音轨的提取

音乐文件格式主要有 3 类: 音频文件、模块文件和 MIDI 文件. 与其他两种格式相比, MIDI 文件具有文件小、可编辑性强、处理速度快以及文件通用性好的特点. 因此很多音乐特征研究的工作都用 MIDI 格式的音乐作为样本库.

大部分的 MIDI 都是多音轨文件, 包括主旋律所在的主音轨以及其他伴奏旋律所在的音轨. 事实上, 我们的工作只分析主旋律中的特征, 而忽略伴奏信息, 所以在 MIDI 各个轨道中提取出主旋律所在的主音轨是 MIDI 音乐情感分析的前提. 首先通过格式转换的相关程序, 将 MIDI 格式的文件转换成文本文件进行保存, 然后使用分类算法提取其中的主音轨. 具体分类算法如下:

首先根据信息熵理论定义音轨特征的熵值, 然后由 MIDI 文件的音轨信息熵和其他重要特征组成特征向量, 构建随机森林分类器抽取 MIDI 文件主旋律所在的音轨. 实验表明, 该模型有较高的准确率, 能有效地提取 MIDI 文件主旋律音轨. 由这种音轨特征向量分别构成的随机森林分类器, 抽取主旋律的正确率可以达到 93.53%[9].

2.2　音乐特征的抽取

音乐是由一连串音符所组成的, 每个音符又包含了诸如音高、时值、力度等信息; 但是音乐情感不是通过单独的音所表现的, 而是通过整体旋律展现的. 除了上面提到的音符的一些特性外, 旋律的进行速度、调式等等都会对音乐情感的表现产生重要的影响. 因此, 抽取音乐基本特征的时候必须是乐曲的整体特征. 本文所选择的音乐样本为当代的流行音乐, 由于流行音乐对于调式的敏感度不高, 所以在选择特征的时候忽略调式这个因素. 本文中选取以下 6 个的特征值构成音乐的特征空间, 这 6 个特征值从完成预处理的 MIDI 文件当中抽取:

1)　节拍: 是音乐中有规律地强拍和弱拍的反复, 节拍是音乐的骨架, 同速度特征一样, 节拍特征值也可以直接从 MIDI 文件中读取, 用 B 表示;

2)　变化音的个数: 所谓的变化音就是把固定的音升高或者降低, 在 MIDI 文件中有相应的变音记号来表示, 变化音会对乐曲造成冲突不和谐的感觉, 对乐曲的情感变化有一定影响作用, 在文中用表示 N_{ch};

3)　最大音程: 表示在整首乐曲中, 音高最高的音符与音高最低的音符之间的音程差, 用 PI 来表示音高 pitch, 于是最大音程可表示为 $In = PI_{\mathrm{high}} - PI_{\mathrm{low}}$;

4)　音符密度: 即平均每小节包含的音符数, $D = \frac{N_{PI}}{N_{\mathrm{bar}}}$, 其中规范后的所有样本在每个小节的结尾处都有标记, 而 MIDI 转换之后的文本文件中除首行以外其他每一行都代表一个音符, 于是很容易统计音符数以及小节数;

5)　速度: 每个 MIDI 文件都会记录着一首曲子的速度, 因此这一特征直接从 MIDI 文件中读取即可, 用 V 来表示, 单位为每秒钟的音符数(Note/s);

6)　大和弦小节的比例: 大小三和弦在音乐中对情感有着非常重要的影响作用, 一般都认为大三和弦色彩明亮, 而小三和弦情感色彩相对暗淡; 我们将与大三和弦关系紧密的小节称为大和弦小节, 反之则为小和弦小节, 计算每一首乐曲中所有小节的大小和弦小节的比例, 用 R 表示.

于是, 对于任意一首乐曲(篇幅控制在一定长度之内的作品), 都可以找到一个 6 维的向量来表示其特

征, 其中 $\boldsymbol{F}=[D, B, N_{ch}, In, V, R]$.

3 基于 GEP 的情感识别方法

在抽取出音乐特征并且确定了情感标注之后, 下一步就是构建情感识别系统. 情感判别的任务是进行情感相似性判断. 系统根据情感标注后的样本通过某种学习策略找到音乐情感识别的规律性, 从而建立认知判别公式. 在系统遇到新音乐的时候, 根据总结的公式确定音乐情感向量. 文献[8] 曾经将 GEP 用于情感识别的工作, 与其他机器学习的方法相比, 有非线性表达能力更强, 算法速度更快等优点. 因此我们采用 GEP 算法进行我们的工作, 但是在特征值以及音乐情感的表示都与文献[8] 有很大不同.

3.1 GEP 算法[10]

GEP 是一种新颖的遗传算法, GEP 结合了遗传编程(GP) 和遗传算法(GA) 的优点, 可以用简单的编码解决复杂问题.

GEP 处理的对象可以是单基因或者多基因组成的染色体. 一定结构的字符串作为遗传物质, 称为基因表达式. GEP 的遗传编码是等长的线性符号串, 称为 GEP 染色体. 一个染色体可以由多个基因组成. 每个 GEP 基因都是由头部和尾部组成, 头部可以包含终结符和函数符号, 而尾部只能包含终结符. 终结符是指程序中的输入、常量以及没有参数的函数. 函数符号可以是相关问题领域中的运算符号, 也可以是程序设计中的一个程序构件. 头部的长度 h 通常依具体问题而定, 而尾部的长度则由以下公式得到: $t=h(n-1)+1$, 其中 n 表示所使用的函数集中需要变量最多的函数的参数个数. GEP 的染色体可以包含一个或多个基因. 多基因染色体中的每个基因都有相同的长度, 分别可以描述为一棵表达式树.

GEP 算法中的遗传算子主要包括以下几种:

1) 选择算子: GEP 中采用 GA 中常用的选择算子, 包括竞标赛选择等.

2) 复制算子: 把选择的个体直接复制到下一代中.

3) 变异算子: 可以作用在染色体的任意位置. 要注意的是, 如果把函数变异为一个终结符, 或者把只有一个参数的函数变异为有两个参数的函数, 则对应的表达式树会发生改变.

4) 插串算子: 随机在基因中选择一个子串, 把它插入到基因头部的任意位置(第一个位置除外) , 头部的符号依次向后挪动, 超过头部长度的编码被丢弃.

5) 根插串算子: 根插串与插串相似, 只是它指定了子串插入的位置只能是头部的第一个位置, 因此要求插入的子串必须以函数开头. 根插入算子首先从头部的任意位置开始向后扫描, 若找到第一个函数, 则以该位置为起始选择一段子串, 然后将该子串插入到头部的第一个位置. 头部原来的符号依次向后挪动, 超过头部长度的部分被丢弃. 若没有扫描到函数, 则什么也不做.

6) 单点重组算子: 对于两个染色体, 随机选择一个交换点, 然后互换交换点后面的染色体部分.

7) 双点重组算子: 对于两个染色体, 随机选择两个交换点, 然后把两个交换点之间的子串互换.

8) 基因重组算子: 该算子只作用于多基因的染色体. 对于两个多基因染色体, 随机选择一个基因, 然后互换两个染色体的相应基因.

对于 GEP 来说, 染色体的适应度评估函数的设计至关重要. 通常可以通过该染色体所代表的表达式计算得到的数据与训练数据的吻合程度来评价.

3.2 基于 GEP 的音乐情感识别算法

本文实现的 GEP 算法采用以下参数设置: 每个染色体包括 10 个基因, 基因头长度 15, 每一代包括 50 个个体; 用到的算子有变异算子、插串和重组算子, 其中, 3 个插串算子概率为 10%, 3 个重组算子概率为 30% , 变异算子按照染色体所含基因的多少决定变异基因位个数; 用到的公式包括四则运算、三角函数, 指数函数幂函数等; 适应度评估函数如下:

$$\mathrm{fitness}(i)=\frac{1}{\frac{1}{n}\sum_{j=1}^{n}(X_{ij}-X_j)^2+1},$$

其中, i 表示第 i 个个体解码得到的公式, j 表示第 j 个样本乐曲, n 为样本个数, X_j 为样本实际的 PAD 分量值, X_{ij} 为用公式计算得到的 PAD 分量值. 具体算法描述如下:

1) 随机创建初始种群;

2) 解码每个个体, 得到相应的审美评价公式;

3) 根据审美评价公式计算样本乐曲的 PAD 分量;

4) 适应度评估;

5) 如果当前代数是否达到预设最大演化代数, 则输出最优个体;

6) 保留最优个体;

7) 采用轮盘算法选择个体;

8) 按一定概率进行复制、变异、插串和重组操作,

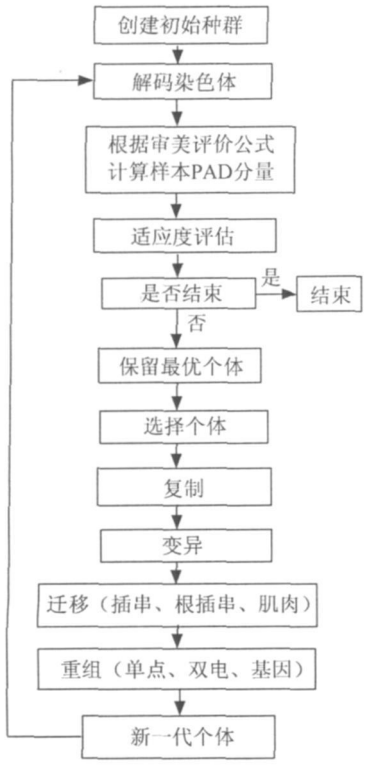

图 1　GEP 算法流程图

Fig. 1　Flow chart of GEP

得到新一代个体;

9) 返回步骤 2) .

整个算法的流程图如图 1 所示.

4　实验结果与分析

我们共收集了 203 首多音轨 MIDI 文件. 它们主要是国内以及港台地区的流行音乐, 经过删选后保留了 145 首. 首先对 MIDI 文件进行预处理, 将 MIDI 中的主音轨识别出来并转换为文本格式; 其次根据我们的要求从这些文本文件当中抽取所需要的 6 个特征值, 得到 145 个 6 维特征向量; 并且由 19 位受访者通过收听 145 首 MIDI 音乐, 给出样本的 PAD 值, 最后计算平均值, 得到每首曲子的 PAD 值. 在这 145 首乐曲中随机挑选其中 25 首作为测试样本, 其余 120 首为 GEP 算法的训练样本. 经过运行, 我们得到以下 3 个公式:

$$P = -0.119 - 2\sin\frac{0.0324}{x_5} + 0.3055x_5 - 2\sin\frac{0.066}{x_1} - \sin\frac{0.066}{x_5},$$

$$A = 0.39x_5 + \frac{x_6}{x_4} - 0.55,$$

$$D = -0.57 + 8.16^{-x_5} + 6.16^{-x_2} + \frac{x_5^{2.22} + x_2^{0.22}}{6.7} + 0.72x_6^{2.18}.$$

为了对算法进行比较, 我们还采用 BP 神经网络对 PAD 建模. BP 神经网络是一种常用的参数建模方法. 我们采用 3 层 BP 神经网络, 每层的神经元数个数分别为 6, 9 和 3. 训练时动量因子设为 0.6, 学习速度为 0.5, 训练目标为 0.01, 迭代次数为 10 000, 并且采用 tansig 函数作为连接函数.

用 GEP 得到的公式和训练得到的 BP 神经网络来计算 25 个测试样本的 PAD 值. 计算得到的 PAD 值与人工标注的 PAD 值进行比较, 分别用公式 $\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2$ 以及 $\frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{x}_i|$ 来计算它们之间的误差, 结果如表 1 所示. 从表 1 中可以看出, 实验结果表明 GEP 算法的误差被控制在较小的范围之内, 但 BP 神经网络的误差比 GEP 算法略高. 另外, BP 神经网络隐藏了音乐的各个变量与 PAD 值的关系, 而通过 GEP 算法得到的公式, 可以更为直观地发现不同的变量如何影响 PAD 取值. 比如, 从 3 个公式中可以看出, 特征量并没有出现在任何一个公式内, 由此也可以得出结论, 变化音的个数这一个特征量还不足以影响整首曲子的情感值. 因此, 我们认为用 GEP 算法更加适合于音乐情感的自动识别.

表 1　GEP 和 BP 算法的 PAD 值误差

Tab. 1　Errors of values of PAD of GEP and BP

观测维度	$\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2$		$\frac{1}{n}\sum_{i=1}^{n}\lvert x_i - \hat{x}_i\rvert$	
	GEP	BP	GEP	BP
P	0.0129	0.0153	0.0654	0.0941
A	0.0074	0.0134	0.0459	0.0854
D	0.0128	0.0229	0.0616	0.1171

GEP 算法以演化代数来控制优化, 演化代数不同对于误差值有一定影响. 图 2 显示了演化代数与计算得到的 PAD 误差的关系. 从图 2 中我们可以看出, 随着演化代数的增加, PAD 值的误差越来越小, 从开始到 100 代, 误差减小较快. 100 代以后, 误差值会继续减少, 但是速度比之前要减缓. 当演化代数达到 500 代

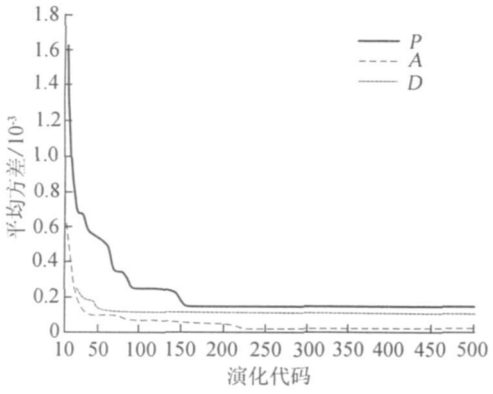

图 2 均方误差变化曲线

Fig. 2 Curve of mean square error

以后,误差达到一个极小值,此后基本保持不变.

5 结 论

情感自动识别对于计算机作曲、音乐检索等都有重要的意义. 本文首先用 PAD 模型来描述音乐作品中的情感状态;并通过分析研究选取 6 个与情感关系密切的音乐特征,以此组成乐曲的特征向量;在此基础上使用 GEP 算法得到音乐情感的 PAD 公式. 实验结果表明,此方法对于情感识别有效,误差值能控制在合理范围内. 下一步的工作将主要针对更为复杂的乐曲,尤其是针对具有复合情感的音乐作品,如何进行情感的分类和识别的研究.

参考文献:

[1] Juslin P, Sloboda J. Music and Emotion: theory and research(Series in Affective Science)[M]. Oxford: Oxford University Press, 2001: 100-120.

[2] Umapathy K, Krishnan S, Jimaa S. Multigroup classification of audio signals using time-frequency parameters[J]. IEEE Trans on Multimedia, 2005, 7(2): 308-315.

[3] Liu D, Lu L, Zhang H J. Automatic mood detection from acoustic music data[C]//Proceedings of the 4th International Conference on Music Information Retrieval. Baltimore, Maryland, USA: Johns Hopkins University, 2003: 81-87.

[4] Pao T L, Cheng Y M, Yeh J H, et al. Comparison between weighted D-KNN and other classifiers for music emotion recognition[C]//Proceedings of the 3rd International Conference on Innovative Computing Information and Control. Dalian, China: IEEE, 2008: 530-533.

[5] Sun Shouqian. Study on linguistic computing for music emotion[J]. Journal of Beijing University of Posts and Telecommunications, 2006, 29(Sup. 2): 35-40.

[6] Sun S Q, Liu T, Wang X, et al. Music's affective computing model based on fuzzy logic[C]//Proceedings of the World Congress on Intelligent Control and Automation (WCICA). Dalian, China: IEEE, 2006: 9477-9481.

[7] Mehrabian A. Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament[J]. Current Psychology, 1996, 14(2): 261-292.

[8] 刘涛. 音乐情感认知模型与交互技术研究[D]. 杭州:浙江大学, 2006.

[9] 黄志刚, 周昌乐, 蒋旻隽. MIDI 主旋律音轨的提取[J]. 厦门大学学报:自然科学版, 2010, 49(1): 43-46.

[10] Ferreira C. Gene expression programming: a new adaptive algorithm for solving problems[J]. Complex Systems, 2001, 12(2): 87-129.

Automatic Recognition of Emotion in Music

JIANG Min-jun, ZHOU Chang-le*, HUANG Zhi-gang

(Fujian Key Laboratory of the Brain-like Intelligent Systems, School of Information Science and Technology, Xiamen University, Xiamen 361005, China)

Abstract: Emotion has great impact on music composition. The development of emotion recognition systems for music is very important to the research of computer music. The paper presents a method based on pleasure arousal dominance (PAD) model and genetic expression programming (GEP) algorithm. In the paper, we advance 6 musical features which have close relations with music emotion and use PAD model to denote different kinds of music emotion. By using GEP algorithm, three formulae can be obtained by which the values of PADs in different songs can be calculated respectively. The experiment shows this method yields good results and achieves high accuracy.

Key words: emotion recognition; PAD model; musical features; GEP algorithm