Universe
Scientific Publishing

基于相关关系构建音乐相似性度量模型

潘旭阳　俎毓伟　杨佳鹏
华北理工大学理学院　河北　唐山　063210

摘　要：音乐自古以来就是人类社会的一部分。为了理解音乐在人类集体中所扮演的角色，理解不同类型音乐之间的相互影响，需要开发一种量化音乐发展的模型。在创作新音乐时，许多艺术家为音乐类型的重大转变作出了巨大贡献，在此期间音乐艺术家会被许多因素影响。我们可以通过考虑歌曲音乐特征的相似程度，来捕捉音乐艺术家之间的相互影响，从而判断不同流派的音乐艺术家间是否会相互影响。

由于变量个数较多，维度较高，所以采取低方差滤波的降维手段并结合音乐影响的现实因素提取重要的维度，然后对不同流派的艺术家进行统计，从而构建音乐相似性度量模型，计算不同流派在各个条目中的平均值，进而通过变异系数来衡量相同流派的离散程度，并且计算各个流派的相关系数来判断流派之间的关联性。可判断出是否随着时间的推移可以改变音乐艺术家的音乐风格。

关键词：量化；相似性度量；低方差滤波；相关系数

1 问题介绍

1.1 问题背景

音乐是人类社会的一部分，是文化遗产的重要组成部分。为了理解音乐在人类集体经验中所扮演的角色，有必要发展一种量化音乐发展的方法。影响艺术家创作新音乐的因素有很多，包括其天生的创造力，当前的社会或政治事件，使用新乐器或工具的机会或其他个人经历等。我们的研究目标是理解和衡量先前制作的音乐对新音乐和音乐艺术家的影响。

许多歌曲都具有类似的声音，许多艺术家为音乐类型的重大转变做出了贡献。有时，这些变化是由于艺术家间的影响，有时，这是对外部事件的响应。我们通过考虑歌曲的网络和给定数据集中的音乐特征来捕捉音乐艺术家之间的相互影响。

1.2 问题重述

基于 5026 个影响者，3761 个追随者，15 个音乐特征，20 个流派的详细数据 ，我们需要解决：

➢ 根据相关数据集制定音乐相似度的度量，观察同一类型的艺术家是否比不同类型的艺术家更相似？

2 音乐相似性度量模型的建立

2.1 基于低方差滤波模型进行降维

我们利用 FULL_MUSIC_DATA 和音乐特征的两个数据集来研究音乐的相似性。由于变量条目个数较多，纬度较高，分析起来比较繁琐，有些条目之间存在正相关的关系，有些条目对所研究的问题没有关系或者关系不大。因此，我们需要采用降维的手段提取出几个重要的纬度。

首先，需要对数据进行归一化 [1] 处理。因为已经假设方差小的特征表示该特征的信息量很小，所以采用低方差滤波模型进行降维处理。具体流程如下：

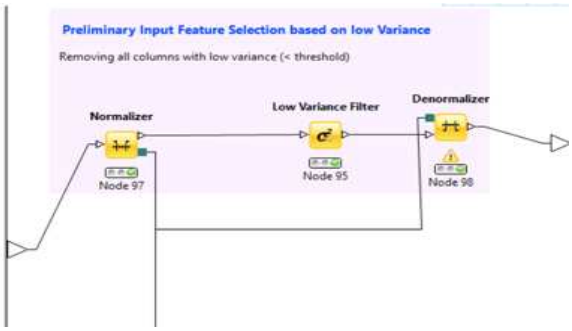

图 1：低方差模型流程图

提取结果如下表 1：

表 1：提取出来的重要维度

Duration_ms	tempo	loudness	mode	key
7.00×10^{9}	3.68×10^{2}	1.75×10^{1}	1.47×10^{1}	1.24×10^{1}

2.2 流派内相似性度量模型的建立

定义：在概率论和统计学中，变异系数 [2] 又称“离散系数”，是概率分布离散程度的一个归一化量度，其定义为标准差与平均值之比。

数据降维处理之后，开始对相似性度量模型进行构建。首先，根据这些艺术家所属的流派进行统计，计算出各个流派在不同特征中的平均值，进而通过计算同一流派内各个特征的变异系数来初步衡

量相同流派的离散程度。

变异系数可以消除单位差异和平均差异对两个或两个以上数据变异程度的影响，也就是去除了量纲的影响。当变异系数值较低时，数据具有较小的变异性和较高的稳定性，如图 2 所示。

通过对图 1 的观察，发现在同一流派内，共有 5 个特征，并没有一个标准来衡量同一流派内的相似度。所以，在图 1 的基础上，对每一流派的各个特征的变异系数取平均值，以此作为衡量相似度的标准，如下图 3 所示。

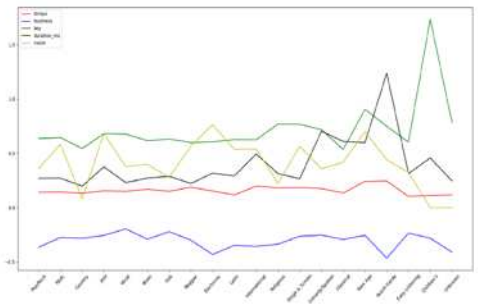

图 2：同一流派内各特征变异系数

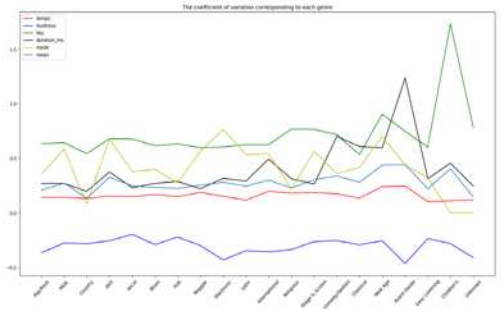

图 3：加入标准值的变异系数

通过观察各个流派内 5 个特征的变异系数的平均值，发现相同流派内的变异系数几乎都达到了中等变异的程度，即相同流派内并非更相似，但有一定聚集的程度。而在后面的一些流派中有些变异系数超过了 40%，其离散程度很大，推断是受其他流派音乐风格较大的影响。

2.3 流派间相似性度量模型的建立

定义：相关关系 [3] 是一种非确定性的关系，相关系数 ε 是研究变量之间线性相关程度的量。

$$\varepsilon = \frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}}$$

其中 $Cov(X,Y)$ 两个变量的协方差，$Var[X]$，$Var[Y]$ 分别为两个变量的标准差。

接下来计算各个特征在不同流派中的平均值，对数据进行归一化处理，进而通过 Python 画出不同流派间相关系数的热力图，如图 4 所示：

通过观察图 4，发现部分流派间的相似程度非常高，例如 Vocal、Blues and Folk。可以很明显的发现不同流派之间会有一定相关性，甚至有些流派几乎达到正负相关，因此可以初步判定流派之间会互相影响，从而随着时间的推移达到改变音乐艺术家的音乐风格。

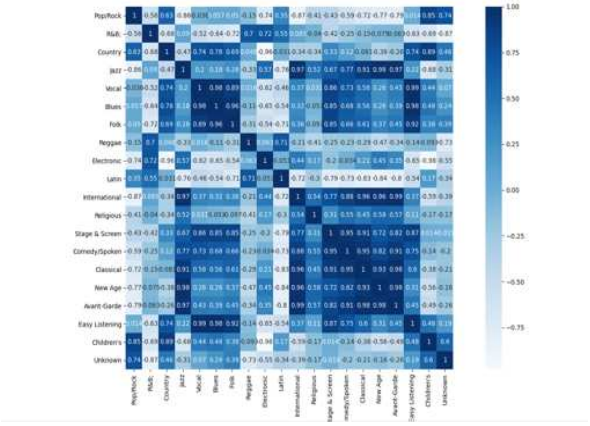

图 4 不同流派间的相关系数

3 结论

我们首先运用低方差滤波对数据进行降维操作，利用降维后得到的数据集对流派间、流派内分别进行相似性度量，所得到结果如下：

1 相同流派内的变异系数较高，即并非相同流派内更相似，但仍有一定的聚集程度。

2 不同流派间会有一定的相关性，甚至有流派几乎达到正负相关，所以可初步判定流派间会相互影响并起到改变音乐艺术家的音乐风格的作用。

3 同一类型的音乐艺术家并不一定比不同类型的音乐艺术家相似。

参考文献：

[1]yehui_qy. 多种数据过滤与降维算法 .https://blog.csdn.net/yehui_qy/article/details/54314795.

[2] 百度百科 . 变异系数 .https://baikes.baidu.com/item/%E5%8F%98%E5%BC%82%E7%B3%BB%E6%95%B0/6463621?fr=aladdin.

[3] 尚轶伦 . 相关系数 .https://baike.baidu.com/item/%E7%9B%B8%E5%85%B3%E7%B3%BB%E6%95%B0/3109424?fr=aladdin. 同济大学数学科科学院 .