2021 ICM D

音乐的影响

音乐自古以来就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，我们被要求开发一种量化音乐演进的方法。当艺术家创作一首新的音乐时，有许多因素可以影响他们，包括他们与生俱来的独创性，当前的社会或政治事件，获得新乐器或工具的机会，或者其他个人经历。我们的目标是了解和衡量以前制作的音乐对新音乐和音乐艺术家的影响。

一些艺术家可以列出十几个甚至更多的影响了他们自己的音乐作品艺术家。也有人建议，影响可以通过歌曲特征(如结构、节奏或歌词)之间的相似程度来衡量。音乐有时会有革命性的变化，以提供新的声音或节奏，例如当新的流派出现时，或者现有流派的再创造(例如古典、流行/摇滚、爵士乐等)。这可以归因于一系列的小变化，艺术家的合作努力，一系列有影响力的艺术家，或者社会内部的转变。

许多歌曲都有相似的旋律，许多艺术家对音乐流派的重大转变做出了贡献。有时，这些变化是由于一位艺术家影响了另一位艺术家。有时，它是对外部事件(如重大世界事件或技术进步)的响应而出现的变化。通过考虑歌曲的网络和它们的音乐特征，我们可以开始捕捉到音乐艺术家对彼此的影响。也许，我们还可以更好地理解音乐是如何随着时间的推移在社会中演变的。

您的团队受综合集体音乐协会(the Integrative Collective Music，ICM)委托来开发一种衡量音乐影响力的模型。这个问题要求你考察艺术家和流派的进化和革命趋势。为此，ICM 向您的团队提供了几个数据集：

（1）“Influence\_Data”（这些数据是从网站 AllMusic.com 收集来的）代表艺人自己报道的音乐影响者和追随者，以及行业专家的意见。这些数据包含 5854 名艺术家在过去 90 年中的影响力和追随者。

（2）“full\_music\_data”(这些数据来自 Spotify 的 API)提供了 16 个变量条目，包括舞蹈性、节奏、响度和音调等音乐特征，以及针对 98,340 首歌曲的 artist\_name 和 artist\_id。这些数据用于创建两个汇总数据集，包括:

a. 艺术家的平均值“data\_by\_artist”，

b. 表示跨年“data\_by\_year”。

注意：这些文件中提供的数据是较大数据集的子集。这些文件包含的数据是您用于求解此问题的唯一数据。

为了执行这一具有挑战性的项目，ICM 协会要求您的团队通过对音乐艺术家的影响来探索音乐的演变，具体做法如下：

（1） 使用 Influence\_Data 数据集或其中的一部分创建音乐影响力的(多个)定向网络，其中影响者连接到追随者。开发参数来捕捉这个网络中的‘music influence’。通过创建定向影响者网络的子网络来探索音乐影响力的子集。描述此子网。在这个子网络中，你的‘music influence’指标揭示了什么？

（2） 使用 FULL\_MUSIC\_DATA 和/或音乐特征的两个汇总数据集(具有艺术家和年份)来开发音乐相似性度量模型。用你的衡量标准，流派内的艺术家比流派间的艺术家更相似吗？

（3） 比较流派之间和流派内部的相似之处和影响。流派的区别是什么？流派是如何随着时间的推移而变化的？有些流派与其他流派有关联吗？

（4） 说明 DATA\_ENAFSONCE 数据集中报告的相似性数据是否表明识别出的影响者实际上影响了各自的艺术家。“有影响力的人”真的会影响追随者创作的音乐吗？是某些音乐特征比其他特征更具“感染力”，还是它们在影响某个特定艺术家的音乐方面都扮演着相似的角色？

（5） 从这些数据中找出音乐演进过程中是否有可能革命性(重大飞跃)的特征？在您的网络中，哪些艺术家代表变革者(重大变革的影响者)？

（6） 分析一种流派中随着时间发生的音乐演变的影响过程。您的团队能否确定能够揭示动态影响者的指标，并解释流派或艺术家如何随着时间的推移而变化？

（7） 你的工作如何表达音乐在时间或环境中对文化的影响？或者，如何在网络中识别社会、政治或技术变化(如互联网)的影响？

请给 ICM 协会写一份一页的文档，说明使用你的方法来理解音乐通过网络的影响的价值。考虑到这两个问题数据集的内容仅限于某些流派，以及随后对这两个数据集共有的艺术家的限制，随着更多或更丰富的数据，您的工作或解决方案将如何改变？建议进一步研究音乐及其对文化的影响。

ICM 协会是一个涉及音乐、历史、社会科学、技术和数学领域的跨学科和多样化的团体，期待着您的最终报告。

您总共不超过 25 页的 PDF 解决方案应包括：

* 单页摘要表
* 目录
* 您的完整解决方案
* 给 ICM 协会的一页文件
* 参考文献列表

本届赛事的新注意事项：ICM 大赛现在有 25 页的限制。您提交的所有内容都计入 25 页的限制：摘要表、目录、解决方案正文、图片和表格、一页文档、参考列表以及任何附录。

附件：

我们为这个问题提供了以下四个数据文件。所提供的数据文件包含您用于解决此问题的唯一数据。

1. influence\_data.csv

2. full\_music\_data.csv

3. data\_by\_artist.csv

4. data\_by\_year.csv

数据说明：

1. influence\_data.csv (数据以 utf-8 编码，以便处理特殊字符):

influencer\_id:给被列为影响者的人的唯一标识号。(数字或者字符串)

influencer\_name:追随者或行业专家给出的有影响艺术家的姓名。(字符串)

influencer\_main\_genre: 最能描述有影响艺术家创作的大部分音乐的流派。(如果有的话)(字符串)

influencer\_active\_start: 这位有影响力的艺术家开始音乐生涯的前十年。(int 形)

follower\_id: 给被列为追随者的艺术家的唯一标识号。(数字或者字符串)

follower\_name:追随有影响力的艺术家的艺术家的名字。(字符串)

follower\_main\_genre: 最能描述以下艺术家创作的大部分音乐的流派。(如果有的话)(字符串) follower\_active\_start: 追随的艺术家音乐生涯的前十年。(整数)

2. full\_music\_data.csv

3. data\_by\_artist.csv

4. data\_by\_year.csv

音乐的特征:

danceability:：根据音乐元素(包括节奏、节奏稳定性、节拍强度和整体规律性)的组合来衡量一条赛道是否适合跳舞的指标。值为 0.为 0 最不适合跳舞，值为 1.0 最适合跳舞。

(浮点数)

energy:一种表示对强度和活动性的感知的量度。值为 0.0 表示最不强烈/充满活力，值为 1.0表示最强烈/充满活力。通常，充满活力的曲目给人的感觉是快速、响亮和嘈杂。例如，死亡金属的能量很高，而巴赫的前奏曲在音阶上得分很低。影响这一属性的感知特征包括动态范围、感知响度、音色、起始率和总熵。

(浮点数)

Valence：描述一首曲目所传达的音乐积极程度的一种衡量标准。值 0.0 表示最负，值 1.0 表示最正。高价的曲目听起来更积极(如高兴、欢快、欢快)，而低价的曲目听起来更消极(如悲伤、沮丧、愤怒)。(浮点数)

tempo: 曲目的总体估计速度，以每分钟节拍(BPM)为单位。在音乐术语中，节奏是给定乐曲的速度或节奏，直接从平均拍子时长得出。(浮点)

loudness: 音轨的整体响度，以分贝(DB)为单位。值的典型范围在-60 到 0 db 之间。响度值是整个音轨的平均值，对于比较音轨的相对响度很有用。响度是声音的性质，它是物理强度(振幅)的主要关联。

mode: 曲目的模态(大调或小调)、旋律内容来源的音阶类型的指示。主要由 1 表示，次要为0。

key:估计的音轨的整体数值。为使用标准音调等级记法映射到音调的整数。例如，0=C， 1=C♯/D♭，2=D，依此类推。如果未检测到 key 值，则的值设定为-1。(整数)

人声类型:

acousticness: 音轨是否是声学的置信度测量(没有技术增强或电子放大)。值 1.0 表示音轨的可信度很高。(浮点型)

instrumentalness: 预测曲目是否不包含人声。在这种情况下，“噢”和“啊”的声音被视为器乐。说唱或有声歌曲显然是“有声的”。器乐度值越接近 1.0，曲目中不包含声音内容的可能性就越大。高于 0.5 的值用于表示乐器音轨，但随着该值接近 1.0，置信度会更高。(浮点型) liveness: 检测曲目中是否有观众。活跃度值越高，表示实时执行曲目的概率越高。如果值大于 0.8，则表示音轨处于活动状态的可能性很大。(浮点数)

speechiness:检测音轨中是否存在口语单词。越是专门的演讲——比如录音(例如脱口秀、有声读物、诗歌)，属性值越接近 1.0。大于 0.66 的值描述可能完全由口语单词组成的曲目。介于 0.33 和 0.66 之间的值描述可能同时包含音乐和语音的曲目(分段或分层)，包括说唱音乐等情况。低于 0.33 的值很可能表示音乐和其他非语音轨道。(浮点数)

explicit:检测曲目中的显式歌词(TRUE(1)=是，是；FALSE(0)=否，不是或未知)。(布尔值)

描述：

duration\_ms: 音轨的持续时间，以毫秒为单位。(整数)

popularity:这条音轨的受欢迎程度。该值将介于 0 和 100 之间，其中 100 是最受欢迎的。人气是通过算法计算的，并且在很大程度上是基于曲目已经播放的总次数和这些播放的最近程度。一般来说，现在播放频率较高的歌曲会比过去播放频率较高的歌曲更受欢迎。

重复的曲目(例如，来自单曲和专辑的相同曲目)是独立评级的。艺术家和专辑的受欢迎程度是从曲目受欢迎程度定量推导出来的。(整数)(布尔值)

year: 发行曲目的年份。(1921 年至 2020 年的整数)

release\_date: 发行曲目的日期大多采用 yyyy-mm-dd 的格式，但是日期的精度可能会有所不同，有些只是以 yyyy 给出。

song\_title (censored): 曲目的名称。(字符串) 歌曲标题中任何潜在的明确单词已经被程序删除。

count:特定艺术家的歌曲数量在 full\_music\_data.csv 文件中表示。

2、3参考之前发的网络入侵

2、音乐相似性度量模型

数据预处理后，KNN聚类

通过点的一簇簇表示相似度（距离远近），标记颜色表示流派。从簇与簇、颜色与颜色之间的关系，直观的看出流派与相似度的关系，流派间相似度

注意：此处流派不能作为因变量，仅用相似度。后续补充颜色表示流派

3、流派的区别

有监督学习

流派画成标记，拟合，得到函数

此处不同颜色区域放大，单独分析。得出的结果再用多维雷达图（音量、曲调等）表示，以此可以相互之间比较

4、音乐特征比其他特征更具“感染力”

判断相似度的热点图

之前默认相似性，这边判别指标

第一问中个体与其他个体的相似度，衡量是否追随，得出追随者关系

在2、3问基础上增加细节，每个方向做一个簇，每个方向进行比较。“有影响力的人”是各个方面都有影响还是单个方面影响度极大

5、音乐演进过程中革命性的特征

某种流派的突然增加或者突然减少，画图分析整体走向。参考各个国家随着时间的动态变化图，整体趋势。 发生重大历史事件的时刻的对照

按照时间划分，画横竖轴

LSTM

6、分析一种流派中随着时间发生的音乐演变的影响过程

在问题5得出流派在各个年代的变化特性。（可能特定定向移动）

指标在演变过程中的变化

7、你的工作如何表达音乐在时间或环境中对文化的影响？

考虑量化指标评估