

2021年ICM

问题D：音乐的影响

自古以来，音乐就已成为人类社会的一部分，已成为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发展的方法。在创作新音乐时，有许多因素会影响艺术家，包括其天赋的创造力，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。我们的目标是了解和衡量先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或更多他们认为对自己的音乐作品有影响的艺术家。还建议可以通过歌曲特征（例如结构，节奏或歌词）之间的相似程度来衡量影响力。音乐有时会发生革命性的变化，提供新的声音或节奏，例如何时出现新的流派，或者对现有流派（例如古典，流行/摇滚，爵士等）进行重新发明。这可能是由于一系列小变化，艺术家的合作努力，一系列有影响力的艺术家或社会内部的变化所致。

许多歌曲具有相似的声音，许多艺术家为音乐类型的重大转变做出了贡献。有时，这些变化是由于一位艺术家影响了另一位艺术家。有时，这是对外部事件（例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及其音乐特征，我们可以开始捕捉音乐艺术家之间的相互影响。而且，也许，我们还可以更好地了解音乐随着时间的流逝在整个社会中的发展。

集成集体音乐（ICM）协会确定了您的团队，以开发一种衡量音乐影响力的模型。这个问题要求您检查艺术家和流派的进化和革命趋势。为此，ICM为您的团队提供了一些数据集：

- 1) “influence_data”¹代表艺术家自己报告的音乐影响者和追随者，以及行业专家的意见。这些数据包含过去90年中5,854位艺术家的影响者和关注者。
- 2) “full_music_data”²提供了16个变量项，包括音乐特征（如舞蹈性，速度，响度和调子），以及98,340首歌曲中的每一个的artist_name和artist_id。这些数据用于创建两个摘要数据集，包括：
 - a. 艺术家“data_by_artist”的平均值，
 - b. 表示跨年“data_by_year”。

¹ 这些数据来自AllMusic.com

² 这些数据来自Spotify的API

注意：这些文件中提供的数据是较大数据集的子集。这些文件包含您应为该问题使用的唯一数据。

为了执行这个具有挑战性的项目，ICM协会要求您的团队通过以下措施，通过音乐艺术家随时间的影响来探索音乐的发展：

- 使用Impact_data数据集或其中的一部分来创建音乐影响力的（多个）定向网络，将影响者连接到追随者。开发可捕获此网络中“音乐影响力”的参数。通过创建一个直接影响者网络的子网络来探索音乐影响力的子集。描述这个子网。你的“音乐影响”措施在这个子网中显示了什么？
- 使用音乐特征的full_music_data和/或两个音乐特征摘要数据集（包括艺术家和年份）来制定音乐相似度的度量。用你的方法，同一类型的艺术家比不同类型的艺术家更相似吗？
- 比较流派之间和流派之间的相似性和影响。什么是流派的区别，流派如何随时间变化？有些类型与其他类型有关吗？
- 指出数据影响数据集中所报告的相似数据是否暗示已识别的影响者实际上影响了各自的艺术家。“影响者”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“具有传染性”，还是它们在影响某一特定艺术家的音乐方面都有相似的作用？
- 从这些数据中找出是否有可能意味着音乐进化中的革命(重大飞跃)的特征？在你的网络中，哪些艺术家代表革命者(影响专业变化的人)？
- 分析音乐演变的影响过程。你的团队能识别出反映动态影响因素的指标，并解释音乐类型或艺术家是如何随着时间的推移而变化的吗？
- 您的作品如何表达关于音乐在时间或环境中的文化影响的信息？或者，如何在网络中识别社会、政治或技术变革(如互联网)的影响？

向ICM协会写一份一页纸的文件，说明使用您的方法通过网络理解音乐影响的价值。考虑到这两个问题数据集仅限于某些类型，然后又针对这两个数据集共有的艺术家，您的作品或解决方案将如何随着更多或更丰富的数据而发生变化？建议进一步研究音乐及其对文化的影响。

来自音乐，历史，社会科学，技术和数学领域的跨学科，多元化的ICM协会期待您的最终报告。

您的PDF解决方案（总共不超过25页）应包括：

- 一页的摘要表。
- 目录。
- 您的完整解决方案。
- 对ICM协会的一页文件。
- 参考文献清单。

注意：2021年的新功能！现在，ICM竞赛限制为25页。提交的所有方面均计为25页的限制：摘要表，目录，解决方案主体，图像和表格，一页文档，参考列表和任何附录。

附件

针对此问题，我们提供了以下四个数据文件。提供的数据文件包含您应用于此问题的唯一数据。

1. [impact_data.csv](#)
2. [完整音乐数据.csv](#)
3. [数据_by_artist.csv](#)
4. [data_by_year.c](#)

sv数据描述

1. [impact_data.csv](#)

（数据以utf-8编码，以允许处理特殊字符）：

- `impactr_id`: 给列出为影响者的人的唯一标识号。（数字字符串）
- `impactr_name`: 影响者的名称，由关注者或行业专家提供。（串）
- `impactr_main_genre`: 最能描述有影响力的艺术家创作的大部分音乐的流派。（如果有）（字符串）
- `impactr_active_start`: 影响力艺术家开始其音乐事业的十年。（整数）
- `follower_id`: 提供给列出为关注者的艺术家的唯一标识号。（数字字符串）
- `follower_name`: 跟随有影响力的艺术家的艺术家的名字。（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）（字符串）
- `follower_active_start`: 以下艺术家开始音乐生涯的十年。（整数）

2. 完整音乐数据.csv 三。数据_by_artist.csv 4数据按年份.csv

从“full_music_data”，“data_by_artist”，“data_by_year”中发现Spotify音频功能：

- **artist_name**: 执行曲目的艺术家。（数组）
- **artist_id**: 受影响的data.csv文件中提供的相同唯一标识号。（数字字符串）

音乐的特点：

- **舞蹈性**: 根据节奏，节奏稳定性，拍子强度和整体规律性等音乐元素的组合来衡量轨道适合跳舞的方式。值0.0最低可跳舞，而1.0最高可跳舞。（浮动）
- **能源**: 表示对强度和活动的感知的量度。值0.0最小强度/能量，而1.0强度最大/能量。通常，充满活力的曲目会感觉快速，响亮且嘈杂。例如，死亡金属具有较高的能量，而巴赫前奏的得分则较低。有助于此属性的感知特征包括动态范围，感知的响度，音色，发作率和一般熵。（浮动）
- **价**: 一种描述曲目传达的音乐积极性的量度。值0.0最负，1.0最正。价态高的音轨听起来更积极（例如，快乐，开朗，欣快），而价态低的音轨听起来更负面（例如，悲伤，沮丧，愤怒）。（浮动）
- **速度**: 曲目的总体估计拍速，以每分钟拍数（BPM）为单位。用音乐术语来说，节奏是指给定乐曲的速度或节奏，它直接来自平均拍子持续时间。（浮动）
- **响度**: 轨道的整体响度，以分贝（dB）为单位。值的典型范围是-60至0 db。响度值是整个轨道的平均值，可用于比较轨道的相对响度。响度是声音的质量，它是身体力量（振幅）的主要心理关联。（浮动）
- **模式**: 音轨的模式（主要或次要）的指示，其旋律内容所源自的音阶类型。Major用1表示，minor用0表示。
- **键**: 曲目的估计总体密钥。整数使用标准音高类别符号映射到音高。例如。0 = C，1 = C# / D b，2 = D，依此类推。如果未检测到密钥，则密钥的值为-1。（整数）

人声类型：

- **声音**: 磁道是否是声学的置信度量（不增强技术或增强电性能）。值1.0表示轨道是声学的高置信度。（浮动）
- **工具性**: 预测曲目是否不包含人声。在这种情况下，“哦”和“啊”的声音被视为乐器。说唱或口语单词轨迹显然是“声音”。器乐性值越接近1.0，则曲目中没有入声内容的可能性越大。高于0.5的值旨在表示乐器音轨，但随着该值接近1.0，置信度更高。（浮动）
- **活跃度**: 检测曲目中观众的存在。较高的活跃度值表示增加了实时执行轨道的可能性。高于0.8的值很可能会显示该轨道处于活动状态。（浮动）

- **言语能力：**检测曲目中口语的存在。与录音类似的语音内容（例如脱口秀，有声读物，诗歌）越多，属性值就越接近1.0。大于0.66的值描述的曲目可能完全由口语组成。介于0.33到0.66之间的值描述了可能同时包含音乐和语音的曲目，无论是分段还是分层的（包括说唱音乐）。低于0.33的值最有可能代表音乐和其他非语音类曲目。（浮动）
- **明确的：**检测曲目中的显式歌词（true（1）=是； false（0）= no，它不是OR未知）。（布尔值）

描述：

- **duration_ms：**轨道的持续时间（以毫秒为单位）。（整数）
- **人气：**这首歌的受欢迎程度。该值将在0到100之间，其中100是最受欢迎的值。受欢迎程度是通过算法计算的，并且在很大程度上取决于音轨的总播放次数以及这些播放的最近时间。一般而言，现在播放频率更高的歌曲将比过去播放频率更高的歌曲具有更高的知名度。重复曲目（例如，同一首曲目和一张专辑中的同一曲目）将被独立评估。艺术家和专辑的流行度是从曲目流行度中数学得出的。（整数）
- **年：**发行曲目的年份。（1921年至2020年的整数）
- **release_date：**曲目发布的日历日期，大多数采用yyyy-mm-dd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（已检查）：**曲目的名称。（字符串）已运行软件以删除歌曲标题中的任何潜在显式单词。
- **计数：**full_music_data.csv文件中表示特定艺术家的歌曲数。（整数）