

D 题翻译：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转变。

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，这种变化是对外部事件（例如重大世界事件或技术进步）的响应而出现的。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 代表音乐人的影响者和追随者，由艺术家本人以及行业专家的意见所报告。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 16 个可变项，包括音乐特征，例如可跳舞性，速度，响度和调子，以及 98,340 首歌曲中的每首的 artist_name 和 artist_id。这些数据用于创建两个摘要数据集，包括：

欢迎加入 2021 美赛跟进群：629043705

a. 艺术家“data_by_artist”的平均值

b. 表示跨年“data_by_year”

注意：这些文件中提供的 DATA 是较大数据集的子集。

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：

(1) 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。通过创建定向影响者网络的子网来探索音乐影响力的子集。描述此子网。您的“音乐影响力”措施在此子网络中体现了什么？

(2) 使用 full_music_data 和/或两个具有音乐特征的摘要数据集（包括艺术家和年份）来制定音乐相似度的度量。使用您的度量，流派的艺术家是否比流派的艺术家更相似？

(3) 比较不同类型之间和内部的相似性和影响。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？

(4) 指出 data_influence 数据集中报告的相似性数据是否表明所确定的影响者实际上会影响相应的艺术家。“影响者”实际上会影响追随者创作的音乐吗？是某些音乐特征比其他音乐特征更具“感染力”，或者它们在影响特定艺术家的音乐方面起着相似的作用？

(5) 从这些数据中确定是否有某些特征可以指示音乐发展中的革命（重大飞跃）？在您的网络中，哪些艺术家代表着革命者（重大变革的影响者）？

(6) 分析一种类型音乐随时间变化的影响过程。您的团队能否确定能揭示动态影响者的指标，并解释流派或艺术家随时间的变化？

(7) 您的作品如何表达有关音乐在时间或环境方面的文化影响的信息？或者，如何在网络中识别社会，政治或技术变化（例如互联网）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

欢迎加入 2021 美赛跟进群：629043705

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 解决方案总页数不超过 25 个，应包括：

- 一页摘要表。
- 目录。
- 完整的解决方案。
- 对 ICM 协会的一页文件。。
- 引用列表。

注意：MCM 竞赛现在限制为 25 页。您提交的所有内容均计入 25 页的限制（摘要表，目录，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. **influence_data.csv**
2. **full_music_data.csv**
3. **data_by_artist.csv**
4. **data_by_year.csv**

资料说明

1. **impact_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。（数字字符串）

欢迎加入 2021 美赛跟进群：629043705

- influencer_name: 影响者的名称，由关注者或行业专家给出。（字符串）
- influencer_main_genre: 最能描述影响艺术家创作的音乐的流派。（如果有）（字符串）
- influencer_active_start: 有影响力的艺术家开始其音乐生涯的十年。（整数）
- follower_id: 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称。（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）（字符串）
- follower_active_start: 以下艺术家开始其音乐生涯的十年。（整数）

2. full_music_data.csv 3. data_by_artist.csv 4. data_by_year.csv

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

- artist_name: 执行曲目的艺术家。（数组）
- artist_id: 在 impact_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：一种根据节奏，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）

-能量：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：以每分钟节拍数（BPM）为单位的曲目的总体估计速度。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）

欢迎加入 2021 美赛跟进群：629043705

-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：轨道的形式（主要或次要）的指示，表示其旋律内容的音阶类型。主音由 1 表示，副音由 0 表示。

-键：轨道的估计总体键。整数使用标准音高类别符号映射到音高。例如。 0 = C, 1 = C# / D \flat , 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：

-声学：轨道是否声学的置信度度量（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）

-语音性：检测音轨中是否存在说出的单词。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）

-显式：检测曲目中的显式歌词（真（1）=是；假（0）=否，它不是或未知）。（布尔值）

说明：

-duration_ms：轨道的持续时间（以毫秒为单位）。（整数）

-流行度：曲目的流行度。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

欢迎加入 2021 美赛跟进群：629043705

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：曲目发行的年份。（从 1921 年到 2020 年的整数）

-release_date：曲目发布的日历日期，主要采用 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 文件中代表特定歌手的歌曲数。（整数）