

问题 D:音乐的影响

音乐从一开始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进化的方法。当艺术家创作一段新的音乐时，有许多因素会影响他们，包括他们天生的创造力，当前的社会或政治事件，使用新乐器或工具，或其他个人经历。我们的目标是了解和衡量之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其他艺术家，他们说他们影响了自己的音乐作品。也有人认为，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。有时，音乐也会发生革命性的转变，提供新的声音或节奏，例如当一种新的流派出现，或现有流派的新发明(如古典、流行/摇滚、爵士等)。这可能是由于一系列的小变化，艺术家的合作努力，一系列有影响力的艺术家，或社会内部的转变。

许多歌曲都有相似的声音，许多艺术家为音乐流派的重大转变做出了贡献。有时这些变化是由于一个艺术家影响另一个艺术家。有时是对外部事件(如世界重大事件或技术进步)的反应而产生的变化。通过考虑歌曲网络及其音乐特征，我们可以开始捕捉音乐艺术家对彼此的影响。也许，我们也可以更好地了解音乐是如何随着社会的变迁而演变的。

你的团队已经被综合集体音乐(ICM)协会确定为一个衡量音乐影响力的模型。这个问题要求你审视艺术家和流派的进化和革命趋势。为了做到这一点，你的团队得到了 ICM 提供的几个数据集：

- 1 代表音乐领域的影响者和追随者，这不仅来自艺人自己的报道，也来自行业专家的观点。这些数据包含了过去 90 年里 5854 位艺术家的影响者和追随者。
- 2) “full_music_data”² 为 98,340 首歌曲提供 16 个变量条目，包括舞蹈性、节奏、响度和键等音乐功能，以及每种歌曲的 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 “data_by_artist” 的平均值，
 - b. 表示年份 “data_by_year” 。

¹ 这些数据是从 AllMusic.com 上搜集来的

² 这些数据来自 Spotify 的 API

注意:这些文件中提供的数据是更大数据集的子集。这些文件包含您应该用于解决这个问题的唯一数据。

为了完成这个具有挑战性的项目，ICM 协会要求您的团队通过以下方式来探索音乐的演变，通过对音乐艺术家的影响来进行研究：

使用 `influence_data` 数据集或其部分来创建一个(多个)定向的音乐影响力网络，其中影响者与追随者相连。开发在这个网络中捕捉“音乐影响”的参数。通过创建直接影响者网络的子网络来探索音乐影响的子网络。描述这个子网。你的“音乐影响”指标在这个子网中揭示了什么？

使用 `full_music_data` 和/或音乐特征的两个总结数据集(包含艺术家和年份)来开发音乐相似性度量。使用你的衡量标准，同一类型的艺术家是否比不同类型的艺术家更相似？

比较不同类型之间的相似之处和影响。类型的区别是什么？类型是如何随着时间变化的？有些类型与其他类型有关联吗？

说明 `data_influence` 数据集中报告的相似度数据是否表明所识别的影响者实际上会影响各自的艺术家。这些“影响者”真的会影响粉丝创作的音乐吗？是某些音乐特征比其他特征更具有“感染力”，还是它们在影响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标志着音乐进化的革命(重大飞跃)的特征？哪些艺术家代表了你网络中的革命者(重大变化的影响者)？

分析一种音乐流派随时间而发生的音乐演变的影响过程。你的团队能否识别出能够揭示动态影响因素的指标，并解释类型或艺术家是如何随着时间而改变的？

你的作品如何表达音乐在时间或环境中的文化影响？另外，如何在网络内识别社会、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，说明使用你的方法通过网络了解音乐影响的价值。考虑到这两个问题数据集仅局限于某些类型，随后又局限于这两个数据集所共有的艺术家，你的工作或解决方案将如何随着更多或更丰富的数据而改变？建议进一步研究音乐及其对文化的影响。

ICM 协会，一个来自音乐，历史，社会科学，技术和数学领域的跨学科和多样化的团体，期待你的最终报告。

不超过 25 页的 PDF 解决方案应该包括: 一页摘要表。

目录。

您的完整解决方案。

给 ICM 协会的一页文件。

引用列表。

注意:新的 2021 年 ICM 竞赛现在有了 25 页的限制。你提交的所有方面都在 25 页的限制之内:摘要表, 目录, 解决方案的主体, 图像和表格, 一页的文档, 参考列表, 和任何附录。

附件

针对这个问题, 我们提供了以下四个数据文件。所提供的文件包含您应该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](#)

数据描述

1. [influence_data.csv](#)

(数据以 utf-8 编码, 以允许处理特殊字符):

- influencer_id: 一个唯一的识别号码, 给人列出的影响者。(数字的字符串)
- influencer_name: 追随者或行业专家给出的具有影响力的艺术家的名称。(字符串)
- influencer_main_genre: 最能描述受影响艺术家创作的大部分音乐的类型。(如果可用)(字符串)
- influencer_active_start: 有影响力的艺术家开始他们的音乐生涯的十年。(整数)
- follower_id: 一个唯一的识别号码给艺术家列为追随者。(数字的字符串)
- follower_name: 受影响的艺术家的名称。(字符串)
- follower_main_genre: 最能描述以下艺术家创作的大部分音乐的类型。(如果可用)(字符串)
- follower_active_start: 以下艺术家开始音乐生涯的十年。(整数)

2.full_music_data.csv

3.data_by_artist.csv

4.data_by_year.csv

Spotify 音频功能从 “full_music_data”，“data_by_artist”，“data_by_year” :-
artist_name:艺术家谁执行的轨道。(数组)

—artist_id:与 “influence_data.csv” 文件中给出的唯一标识号相同。(字符串
的位数)

音乐特点:

-舞蹈性:根据音乐元素的组合,包括节奏、节奏稳定性、拍子力度和整体规律性,来衡量一个曲目是否适合跳舞。0.0 是最不适合跳舞的值,1.0 是最适合跳舞的值。(浮动)

-能量:表示强度和活动的量度。0.0 是强度最小的值,1.0 是强度最大的值。通常,充满能量的轨道给人的感觉是快速、响亮和嘈杂的。例如,死亡金属的能量很高,而巴赫的序曲在音阶上得分较低。对这一属性有贡献的感知特征包括动态范围、感知到的响度、音色、发作率和一般熵。(浮动)

效价:描述音轨所传达的乐感的一种量度。0.0 是最负的值,1.0 是最正的值。高效价音轨听起来更积极(如快乐、快乐、欣快),而低效价音轨听起来更消极(如悲伤、抑郁、愤怒)。(浮动)

-节奏:在每分钟节拍(BPM)的轨道的总体估计速度。在音乐术语中,节奏是一个给定的作品的速度或节奏,直接来源于平均拍子持续时间。(浮动)

—响度:音轨的整体响度,单位为分贝(dB)。数值的典型范围在-60 到 0 分贝之间。响度值平均横跨整个轨道,是有用的比较轨道的相对响度。响度是声音的基本质量
身体力量的心理关联(振幅)。(浮动)

调式:音轨中调式(大调或小调)的一种表示,调式是音阶的一种类型,它的旋律内容来源于此。
大调用 1 表示,小调用 0 表示。

-关键:估计的整体关键的轨道。整数使用标准音高类表示法映射到音高。如 0 = C、1 = C 大调/b, 2 = D 等。如果未检测到键值,则键值为-1。(整数)

类型的声乐:-

音色:衡量音轨是否具有音色(不含技术)的可信指标

增强或电放大)。1.0 代表音轨的高度置信度。(浮动)

-器乐性:预测音轨是否不包含人声。“Ooh”和“aah”的声音在这个语境中被视为有用的。说唱或口语的音轨是明显的“声音”。工具性值越接近 1.0,这首歌就越有可能不包含声乐内容。高于 0.5 的值表示工具性轨迹,但当值接近 1.0 时,置信度更高。(浮动)

活跃度:检测轨道中是否有观众。较高的活跃度值表示实时执行跟踪的可能性增加。如果该值高于 0.8,则表明该轨道很有可能是实时的。(浮动)

-

- **语音:**检测语音在音轨中是否存在。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。高于 0.66 的值表示可能完全由口语单词组成的音轨。0.33 到 0.66 之间的值描述了既包含音乐又包含语言的音轨，可以是分段的，也可以是分层的，包括说唱音乐这样的情况。低于 0.33 的值很可能代表音乐和其他非语言类音轨。(浮动)
explicit:检测音轨中明确的歌词(true (1) = yes;false(0) =不存在或未知)。(布尔)
-

描述:

—duration_ms:磁道持续时间，单位为毫秒。(整数)

- 人气:该赛道的受欢迎程度。该值将在 0 到 100 之间，其中 100 是最常用的。流行度是通过算法来计算的，在很大程度上是基于这首歌的总播放次数以及这些播放的最近时间。一般来说，现在播放频率更高的歌曲会比过去播放频率更高的歌曲更受欢迎。重复的音轨(例如，来自一个单曲和专辑的相同的音轨)是独立评级的。艺术家和专辑的流行度是由歌曲的流行度计算出来的。(整数)

- year:单曲发行的年份。(1921 - 2020 的整数)

release_date:音轨发布的日历日期，大多是 yyyy-mm-dd 格式，但日期的精度可能会有所不同，有些只给出 yyyy。

- song_title(审查):音轨的名称。(字符串)软件运行删除任何潜在的显式文字在歌曲的标题。

- count:在 full_music_data.csv 文件中表示某个艺术家的歌曲数量。(整数)