美赛正文

摘要

音乐随着社会变迁而演变。为了理解音乐在人类集体经验中所扮演的角色, Our team was asked to 审视艺术家和流派随社会变迁的进化和革命趋势。

对于问题一和问题二，我们首先利用邻接表构造艺术家音乐影响网络。在网络中以有向路径加权方式建立影响者与追随者的关系，并通过递归遍历、加权比例求和得出每一艺术家的”音乐影响“参数，即“新生代人才影响力”。该参数揭示了艺术家对新人才吸引力和对音乐发展的贡献。为了确定该路径加权具体数值，我们利用修正余弦相似度算法，辅以适当的数据处理作为相似度度量，从而得出每两位艺术家间的相似度关系，并以此作为加权参数。在此基础上，我们得出了同一类型艺术家比不同类型艺术家更相似的结论。

（需要修改）对于问题三，我们利用相似度度量和“新生代人才影响力”得出各流派间、流派内部的相似性和影响关系。对于流派区别，我们按流派取数据平均，将不同流派平均音乐特征作比较，发现（啥区别）。在此基础上，我们将流派平均数据作相似度分析，配合当时音乐史实验证流派关系。最后我们找到相似度变化大的年份进一步分析（啥音乐）起到代表性作用。

对于问题四，我们将每两艺术家间相似度平均和直接影响相似度平均比较得出粉丝创作受影响者影响。进一步的，我们计算各音乐特征的皮尔森相关系数并进行比较，最后做成柱状图表示他们的关系，发现是（怎么样的，是某系才影响还是都影响）。

对于问题五，我们逐年进行相似度计算，对重点变化年份（）进行音乐特征分析，发现（）。然后，我们在重点变化年份寻找音乐影响大的音乐家，将音乐家与年份进行相似度分析，相似度较大的为革命者，例如（谁）。

对于问题六和七，(待写）

keywords：

Contents

1. Introduction
2. Assumptions and Justification(uncountable)
3. Notations
4. Model developments
5. Sensitivity Analysis
6. Strengths and Weaknesses
7. References
8. Appendices(可有可无？)
9. intro

1.1 background

Music is an art form whose medium is sound and silence.[1] In the thousand-year history of mankind, music is an indispensable part of the development of civilization.When artists are creating music, there are many factors that affect them, including their musical creativity, personal experience, current hot news and social status, and even a trivial voice.Sometimes, music also undergoes revolutionary changes, creating new sounds or rhythms, introducing a new genre and trend, or changing the pattern of existing genres.By analyzing song networks and their musical characteristics in different periods, we can begin to capture the influence of music artists on each other, understand the similarities and differences between and within genres, so as to better understand how music has evolved with social and cultural changes of.

1.2 problem restatement

In order to examine artist genres and evolutionary trends, and build a model of musical influence.We have the following problems waitingtobe solved：

1. 建立一个音乐影响力网络，捕捉“音乐印象”参数。
2. 开发音乐相似性度量。
3. 比较流派内部和流派间艺术家关系，并得到流派间关系。
4. 音乐影响者对于追随者的作用。
5. 在网络中找到可能标志音乐进化的革命。
6. 分析一种流派随时间音乐演变过程，并给出解释。
7. 利用模型表达音乐在时间或环境中的文化影响，识别社会政治或技术变革影响。
8. a ssumptions
9. 假设网络内艺术家可以通过结点的传递间接影响其他艺术家（需要重新斟酌表述）
10. 艺术家之间的相似度与他们各自歌曲数无关
11. 假设各艺术家音乐特征只对各自开始活跃年代有唯一贡献
12. 假设某一流派音乐特征与该年代音乐特征最相似，该流派在该年代最流行。
13. Notations
14. Model development

4.1 预处理

1. 由于data\_by\_artist 数据集，data\_by\_year数据集，full\_music\_data数据集中的各评价指标性质不同，量纲和数量级不同，因此我们对数据进行了标准化处理。

4.2音乐影响力网络与相似度

\*音乐影响网络

我们将影响者和追随者当作网络中的结点，结点间的有向连线表示结点间的影响关系，并将关系量化作为连线权值，整体形成一个加权有向图。如图一所示。

【图片1】

在此基础上，为了更具体说明连线权值，我们把下图【2】中关系作为例子。结点A是B的直接影响者，B是CDE的直接影响者，A是CDE的间接影响者。接下来我们对A与B的音乐特征进行分析，若A与B的音乐特征相似，说明B受到A的影响大，则A对CDE的间接影响大如图【2.1】，若A与B音乐特征不相似，说明B受A的影响小，则A对CDE的间接影响小，如图【2.2】。

【图片2.1】【图片2.2】

此外，间接影响应比实际影响小，因此我们引入了衰减因子a!!!，且有通过数据比对最后确定值为0.2。最后，我们将结点间相似度乘上衰减因子作为连线权值，如下图所示。

【加权分析图】

加权图含义分析。

\*艺术家相似性度量

由data by artist数据集，每位艺术家的音乐特征可看作一个多维向量。余弦相似度算法可以通过计算两个[向量](https://baike.baidu.com/item/%E5%90%91%E9%87%8F/1396519)的夹角[余弦](https://baike.baidu.com/item/%E4%BD%99%E5%BC%A6/73670)值来评估他们的[相似度](https://baike.baidu.com/item/%E7%9B%B8%E4%BC%BC%E5%BA%A6/17931411)，故可以使用余弦相似度算法来计算两个艺术家音乐特征的相似性。其计算公式如下。

余弦相似度计算公式

计算公式解释

进一步的，余弦相似度算法算出两个向量间的夹角的余弦值，为了解决向量夹角过小而向量长度差别很大，但由于余弦值差别小而判断为两向量相似的问题，我们引入了修正余弦相似度算法，公式如下

修正余弦相似度公式

修正余弦相似度公式解释。

最后，由于余弦值的取值范围是【-1，1】，因此我们对修正余弦相似度算法得出的数据进行归一化，得到最终的艺术家间相似度度量。

归一化公式

公式说明

\*网络模型补充

在上述基础上，毋庸置疑的是艺术家可以对自己的前辈产生影响，但该部分影响力对音乐发展推动力有限。为了使”音乐影响“参数更具有现实意义，即从发展的角度看，音乐发展需要注入新鲜血液。因此我们将”音乐影响“定义为”新生代人才影响力“，即艺术家”音乐影响“计算只包含同辈艺术家和后辈艺术家。

此外，由于音乐影响网络中含有回路。我们根据假设，在单个艺术家影响力计算中，不会出现影响自身的情况，故在单次影响力计算中将回路断开，如图三所示。

【图三】

\*音乐影响力

综合上述内容，最后我们可以得出影响力的递归关系式，如下公式

递归关系式

关系式解释。

在我们所建立的网络中，“音乐影响”表示“新生代人才影响力”，我们可以以此揭示出流派中有哪些音乐大师、以及音乐大师在各流派中占比。我们选取流派的前五”音乐影响“艺术家作为大师，并分析出前500音乐影响艺术家绘成图像如下方表示。

【图片3.1】 【图片4.1】

\*艺术家相似度比较

我们以pop/rock流派为例子。对流派内每两位艺术家进行修正余弦相似度计算，求和取平均得到pop/rock流派内平均相似度。再对网络中任意两位艺术家进行相似度计算，求和取平均。

公式

说明，最终得到pop/rock流派内艺术家平均相似度和艺术家总体平均相似度，表格如下

【比较数据表格】

分析表格我们不难得出流派内艺术家比不同类型艺术家更相似的结论。

4.3流派分析

在进行不同流派分析中，对于不知流派的艺术家不参与流派分析，即流派“unknown”不参与分析。

\*流派间分析

我们对已预处理的各流派内部所有艺术家各音乐特征分别进行求和取平均，得出各流派的平均音乐特征。对不同流派进行修正余弦相似度计算，最后进行归一化处理。如下表

【流派间相似性表格】

我们对相似度最低的两个流派进行音乐特征的分析，即（which），可以发现（哪个流派哪个比较高哪个比较低）。类似的我们可以发现，（哪个流派和哪个流派相似度最高，达到了多少），所以可以认为这两个流派是有一定关联的。

接下来，我们利用音乐影响网络计算出每个艺术家的“音乐影响”，将各流派内艺术家的音乐影响进行求和得出流派的音乐影响。如下表

【流派音乐影响表格】

从表格中我们可以得出pop/rock R&B 和 country 是“音乐影响”最高的三个流派，其中pop/rock流派“音乐影响”远超其他流派，说明pop/rock流派吸引的新人才最多，最热门，并且发展最好。

\*流派内部分析

我们对各流派内部每两位艺术家进行相似度计算，求和取平均，得出各流派平均内部相似性，如下表

【流派内部相似性表】

从表中我们可以得知，classical stage&screen easy listening是相似度最高的三个流派，说明这几个流派内部不同艺术家的音乐特征相似。类似的，pop/rock R&B international是相似度最低的三个流派，说明这几个流派内部不同艺术家风格多元，音乐特征离散程度高，差别较大。

\*流行流派随时间变化

根据假设，通过data by artist数据集我们可以得到每位艺术家开始活跃的年代，并以此认定各艺术家音乐特征只对各自相应年代有唯一贡献。将每年代（10年）开始活跃各流派内艺术家音乐特征求平均，得到各流派每十年平均音乐特征。接着，我们在data by year数据集中求得所有流派总体每十年平均音乐特征。通过逐年代对各流派平均音乐特征与年代平均音乐特征做相似度分析 ，可以求得每个年代对应相似度最高的流派，根据假设可以认为该流派在该年代中最流行。流行流派随时间变化即如下图所示

【逐年代流行流派图】

我们可以看到，（哪个流派一直流行）。

4.4 影响者影响力分析

在（哪）中我们已经得出了总体平均相似性，在此基础上，我们求出网络中直接影响的艺术家之间的相似度平均值，如下表格

【影响者影响力分析表格】

由直接影响艺术家相似度平均值更大可以得知“影响者”会影响追随者创作的音乐，追随者会学习影响者的音乐风格，他们之间有更多的相似性。

为了进一步确定追随者受影响的音乐特征，我们使用皮尔森相关系数算法，由下公式

【】

皮尔森相关系数解释。

在网络中，我们分别将影响者的13种音乐特征作为X，被影响者对应音乐特征作为Y，得到13组音乐特征的二维数据（X,Y)，然后对每组音乐特征的数据求皮尔森相关系数，得到13个音乐特征的皮尔森相关系数，画出柱状图如下

【皮尔森相关系数柱状图】

分析柱状图我们可以发现，（是都影响还是只有一个影响）。

4.5 音乐革命分析

首先对data by year 数据集音乐特征进行逐年相似度计算，计算出每相邻两年间的音乐相似度，如下表格

【逐年音乐相似度表】

从表格可以得知（第几年）音乐相似度小，说明年该音乐特征变化大，即发生了音乐进化革命。查数据（分析）。为了确定在网络中的音乐革命者。我们在（哪年）找到同时期活跃的艺术家，将音乐革命年的音乐特征与各艺术家进行相似度分析，发现（谁）的相似度数据大，由于（谁）影响力大并且与当年音乐特征相似度高，故（谁）是革命者。

从图中我们可以得知，1929年、1945年、1967年和1981年与前一年的年音乐特征相似度最低，说明某些这些年份音乐特征变化较大，可能发生了音乐变革。接着，我们在这几个变化年份所在年代中，寻找网络中音乐影响大并且与变化年音乐特征相似度高的艺术家。我们可以认为最终找出来的艺术家就是音乐革命者。在此我们以1967年为例子，发现the beatles，the rolling stones 和 David bowie的影响力比较大，并且与当年音乐特征相似度达到0.9，因此我们认为他们三人是1967年音乐革命的革命者。

4.6 pop/rock流派分析

熵权法可以用熵值来判断某个指标的离散程度，其信息熵值越小，指标的离散程度越大，通过计算各音乐特征的熵值，即可求出动态影响指标。

【熵权法公式】

熵权法说明。

我们在full music data数据集中计算出每年pop/rock的平均音乐特征，以10年为一个区间，利用熵权法计算十年内的音乐特征变化权值，音乐特征权值越大说明该音乐特征在十年内的离散程度越大，动态影响性越大。保留每十年的前三熵值，输出表格如下图

【熵权法数据】

从表中我们可以发现（哪几个）熵值在pop/rock流派发展中一直是动态影响因素指标，（接着从原始数据分析音乐特征随年份变化的特点）。

从表中可以看出，pop/rock流派的 valence，liveness和loudness音乐特征在过去70年中一直是处于高熵权值区域内，即波动性较大的动态音乐特征。这些音乐特征即是反应音乐动态变化的指标。将四个音乐特征进行标准化，绘出图像如下图3。可以看出，在40和50年代，pop/rock流派仍在摸索阶段，流派音乐特征变化差别很大，在60年代后以energy和loudness音乐特征为例，他们总体随时间呈上升趋势，说明pop/rock的响度、活跃程度越来越大。

4.7音乐与文化的关系

要想探究音乐与文化的关系，自然要从音乐对文化的影响和音乐对社会变化的反映两方面讨论。下面详细阐述了使用我们小组构建的影响力网络模型对上述两种关系进行分析的主要思路。

4.7.1音乐对文化的影响

为了得到特定年代音乐对文化的主要影响，我们需要在影响力网络的对应年代子网络中进行探究。首先检索出其中影响力较高的几位艺术家（五位/十位），以表格的形式列出。对他们的流派、音乐风格进行分析，结合历史事件便可以知道这个年代音乐对文化的主要影响力体现在哪里。

4.7.2音乐对文化变化的反映

音乐本身在进行不断的自我更迭、进化，这样的过程是缓慢而又稳定的。但在某些情况下，外部的重大事件能够让音乐得到极大的飞跃和提升。为了说明识别外部影响的思路，我们不妨同样以Pop/Rock流派为例进行分析，而对引起所有音乐流派发生变化的外部影响的分析也是同理。

为了得出某一年代发生的音乐变化，我们使用4.6中通过熵权法得到的重点音乐特征进行分析。将这些挑选后的音乐特征的大小和变化趋势综合分析，最终识别出外部影响。譬如说，结合其他音乐的学术分析文章和史实来看，如果音乐特征中\*\*\*有\*\*\*的趋势，则有可能发生了战争或是

（如果字数不够，就以Pop的某一年代为例为例寻找文献分析重大事件与音乐特征的关系，顺便通过文献来水字数）

1. 模型稳定性

我们对建立的主要模型，即音乐影响模型进行稳定性分析。可知我们最初设定的衰减因素alpha为0.2. 为了探究改变衰减因素对模型结果的影响，将衰减因素为0.2时输出前20位“音乐影响”艺术家，接着将衰减因素调整为0.1再输出前二十位“音乐影响”艺术家。如下表格1

由上表格，

1. 模型优点与缺点

\*优点

1 运用了带权有向图清晰地构建了音乐影响网络，并在“音乐影响”参数中考虑到间接影响因素，影响力结果和史实相比中获得很高的准确度。

2 在算法上我们根据具体需求分别使用了余弦相似度算法，皮尔森相关系数和熵权法，丰富了模型的结构，以这些算法为基础原创了大部分的对比评判模型。

3 在数据处理中，我们为了减少算法误差，使用了修正余弦相似度算法，在数据处理中仅通过标准化处理，在减少量纲误差同时尽量保留了原始数据，并且除去假设中的歌曲数，其他数据都用于模型的建立，数据使用达到最大。

\*缺点

1 为了追求数据准确性，我们没有使用主成分分析法将多维数据进行降维并合理舍去数据，导致程序运行时间很大。

2 在音乐变革模型中，我们没有使用full\_music\_data数据集中的歌曲发行日期作为判断音乐变革艺术家标志，而是以十年为单位在艺术家活跃年份中找到音乐变革的艺术家，数据精度有待提高。

信件

在开始向您介绍我们的模型之前，请允许我们向您的协会在综合音乐领域做出的努力表示由衷的敬佩。

在我们的音乐影响力网络模型中，只要得到艺术家之间的影响关系数据集和艺术家的音乐特征数据集，便可以构建出以艺术家为节点，带有综合影响力和影响关系权值的音乐影响力网络。

在音乐影响力网络上，除了能够直接获得每个艺术家的影响力参数，检索音乐影响力关系将不费吹灰之力，关于其子网的研究也只需要其中的部分数据就可以顺利进行。

从得到的音乐影响力网络和所得到的三个数据集出发，我们的模型不仅可以导出时代音乐变革者的排行，还能通过熵权法计算音乐特征的方式降维数据，进一步分析得出音乐对社会的反映。

值得注意的是，尽管我们的模型在13个音乐特征的分析中拥有比较高的效率，但随着音乐特征数量的增长，我们的模型将不可避免地出现运算时间急剧上升的状况。这时除了换用更高运算能力的计算机外，一种可行的办法是使用主成分分析发来对众多音乐特征降维。通过损失一部分信息换取在低配置计算机上流畅运行的可能性。

如果能够得到音乐艺术家与其他领域人物之间的关系数据集，我们将能够建立一个特殊的跨领域子网。有了这个子网，分析音乐对文化的影响将变得更加容易，而不需要仅通过音乐特征的变化进行分析。譬如说，政治家可以与政府政策相关联，而教师则能够与教育倾向存在联系。

此外，由于此次竞赛中现实条件的限制，我们小组已经对部分问题的计算过程进行了不同程度的简化。如果被给予充足的时间和计算机资源，我们的模型得到的结果将可以保留更多原数据集的完整信息，得到更加准确的结果。

我们所有的小组成员十分期待您对我们的模型的看法。