

# 基于K-means方法的音乐风格特征聚类与分析

A Music Style Analysis Approach Based on K-means Algorithm

答辩人： 解铮

指导教师：董小社

# 研究意义及背景

- 音乐风格分析
  - 理解历史文化的发展
  - 推动心理学等学科发展
  - 音乐生成
  - 音乐推荐
- 音乐家风格分析的传统方法多为人为分析总结。
- 机器学习的方法分析：
  - 有监督的分类器（SVM、神经网络、……）
  - 基于频域分析的统计特征（基于音频格式音乐）

# 已有研究 vs. 本文工作

- 已有研究：

- 现有研究忽略了音乐的结构信息（旋律、节奏、……）
- 有监督方法需要标注数据
- 分析方法脱离音乐理论分析

- 因此，本文提出：

- 从**有结构音乐格式**中进行旋律相关的**特征提取方法**
- 使用K-means算法**无监督**地发现音乐风格的方法
- **通过音乐理论证实结果**，克服没有标注数据的难点
- **整理曲库数据**作为分析材料，克服已有曲库数量/格式缺陷

# 曲库收集

- 海顿、莫扎特、贝多芬的钢琴曲， 共计413首
  - 来自CCARH, PeachNote, Kunst der Fuge, ……
- MusicXML格式——文本描述乐谱

Haydn	Mozart	Beethoven
Piano Sonatas	Piano Sonatas	Piano Sonatas
Masses	Piano Duets	Trios for Piano
Piano Pieces	Piano Variations	Piano Duet
Trios for Piano		Concerto for Piano
Piano Duos		Bagatelle for Piano
		Rondo for Piano
		Fantasia for Piano



# 特征提取方法 - 理论背景

- Implication-Realization 模型 (E. Narmour) :
  - 旋律中两个连续的音程组成的二元组中, 包含着对听众产生心理暗示的信息
- 已有研究工作 (Zivic et al.) :
  - 基于旋律的特征能够支持古典音乐年代的分类
- 综上所述, 有理由认为旋律中的音程二元组隐含了音乐风格信息。

# 特征提取方法 - 旋律抽取



(a) 欢乐颂高音谱表第一乐句



(b) 欢乐颂高音谱表第一乐句（旋律抽取后）

B4	B4	C5	D5	D5	C5	A4	B4	A4	G4	G4	A4	B4	B4	A4	A4
71	71	72	74	74	72	69	71	69	67	67	69	71	71	69	69

(c) 旋律数值序列

(a) 乐谱 -> (b) 旋律 -> (c) 线性数值序列

# 特征提取方法 - 特征矩阵

0 +1 +2 0 -2 -3 +2 -2 -2 0 +2 +2 0 -2 0

(a) 音程数值序列，由图 3-4(c)差分得到

(0, +1) (+1, +2) (+2, 0) (0, -2) (-2, -3) (-3, +2) (+2, -2)  
 (-2, -2) (-2, 0) (0, +2) (+2, +2) (+2, 0) (0, -2) (-2, 0)

(b) 音程二元组序列

$(-3, +2) \times 1$   $(-2, -3) \times 1$   $(-2, -2) \times 1$   
 $(-2, 0) \times 1$   $(0, -2) \times 2$   $(0, +1) \times 1$   
 $(0, +2) \times 2$   $(+1, +2) \times 1$   $(+2, -2) \times 1$   
 $(+2, 0) \times 1$   $(+2, +2) \times 1$

(c) 音程二元组计数

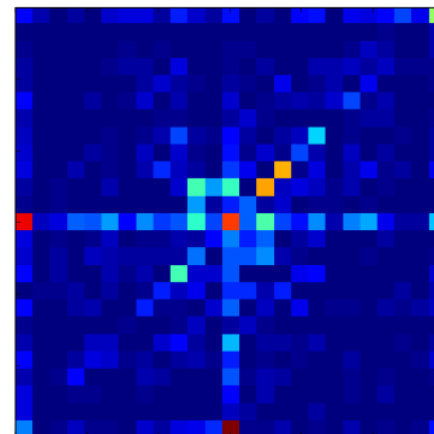
$$\begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \dots & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\ \dots & 1 & 1 & 0 & 2 & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & 0 & 2 & 0 & 0 & 1 & 1 & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\ \dots & 0 & 1 & 0 & 2 & 0 & 1 & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

(d) 特征矩阵，由于旋律中的音程比较小，故数值集中在矩阵中心附近

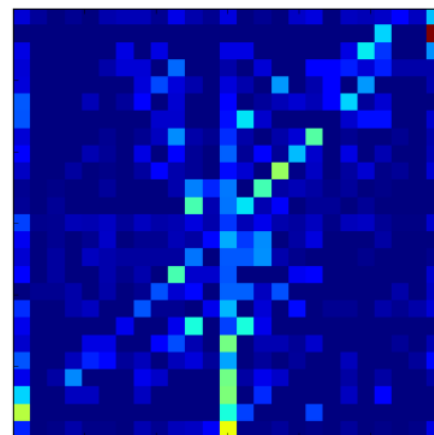
-> (a) 音程数值序列 -> (b) 音程二元组 -> (c) 二元组计数 -> (d) 特征矩阵

# 特征提取方法 - 归一化

条件概率分布  $P_s(i_1, i_2) = \frac{m_{i_1, i_2}^{(s)}}{\sum_{i, j} m_{i, j}^{(s)}}$



联合概率分布  $P_s(i_2|i_1) = \frac{m_{i_1, i_2}^{(s)}}{\sum_j m_{i_1, j}^{(s)}}$



两种矩阵共同生成特征向量



# 聚类分析 - 聚类算法比较

	K-means	K-medians <sup>a</sup>	K-medoids	BIRCH	Linkage	Mean-Shift <sup>b</sup>	AP <sup>c</sup>
<b>#clus.=3</b>							
<i>DBI</i>	3.593	3.679	4.866	<b>3.492</b>	(0.426)	(0.430)	—
<i>DI</i>	0.395	<b>0.469</b>	0.315	0.409	(2.192)	(2.160)	—
<i>SI</i>	<b>0.049</b>	0.022	0.024	0.019	(0.405)	(0.399)	—
MIN/MAX	15.2%/52.2%	23.9%/39.4%	14.7%/45.4%	9.5%/68.8%	0.3%/99.0%	0.3%/99.5%	—
<b>#clus.=4</b>							
<i>DBI</i>	<b>3.315</b>	6.214	4.425	3.486	(0.436)	—	—
<i>DI</i>	<b>0.406</b>	0.236	0.302	0.386	(2.167)	—	—
<i>SI</i>	<b>0.041</b>	0.016	0.019	0.015	(0.393)	—	—
MIN/MAX	11.5%/42.8%	21.2%/27.8%	7.6%/43.6%	8.1%/60.6%	0.3%/98.4%	—	—
<b>#clus.=5</b>							
<i>DBI</i>	<b>2.965</b>	3.934	4.647	3.210	(0.441)	(0.459)	—

\* DBI 越小，或 DI 和 SI 越大，表示分类效果越好。最优的项目表中加粗标注。

\* 有单样例类别 ( $MIN = 0.03\%$ ) 的聚类没有意义，但评估指标会很好。表中用括号标注。

\* DBSCAN 算法在各种参数下都仅能得到一部分数据为噪声，其他全为一类的聚类结果，故未列出。

N

<sup>a</sup> K-medians 在算法中使用曼哈顿距离。为合理评估，指标计算中也使用曼哈顿距离。

<sup>b</sup> Mean-Shift 不接收类别数作为参数，但可以设置带宽参数得到不同的类别数的聚类。

<sup>c</sup> Affinity Propagation 算法可以自行确定类别数。实验中，给出的方案类别数为 11。

# 聚类分析 - 聚类算法比较

	K-means	K-medians <sup>a</sup>	K-medoids	BIRCH	Linkage	Mean-Shift <sup>b</sup>	AP <sup>c</sup>
#clus.=6							
<i>DBI</i>	<b>2.789</b>	6.721	(3.951)	3.636	(0.459)	—	—
<i>DI</i>	<b>0.416</b>	0.208	(0.290)	0.302	(1.968)	—	—
<i>SI</i>	<b>0.042</b>	0.024	(0.019)	0.022	(0.356)	—	—
MIN/MAX	5.2%/38.8%	6.0%/27.3%	0.2%/33.8%	2.1%/46.7%	0.3%/94.2%	—	—
#clus.=7							
<i>DBI</i>	<b>2.424</b>	5.384	(3.782)	3.410	(0.819)	(0.471)	—
<i>DI</i>	<b>0.436</b>	0.244	(0.291)	0.301	(1.100)	(1.888)	—
<i>SI</i>	<b>0.049</b>	0.008	(0.019)	0.023	(0.342)	(0.337)	—
MIN/MAX	1.8%/36.7%	3.9%/25.2%	0.2%/33.6%	1.6%/45.1%	0.3%/94.2%	0.3%/98.4%	—
#clus.=11							
<i>DBI</i>	(2.077)	<b>4.800</b>	(3.128)	(2.568)	(0.687)	—	(2.473)
<i>DI</i>	(0.401)	<b>0.284</b>	(0.275)	(0.334)	(1.229)	—	(0.329)
<i>SI</i>	(0.049)	<b>0.000</b>	(0.018)	(0.018)	(0.326)	—	(0.034)
MIN/MAX	0.3%/38.1%	0.5%/19.4%	0.2%/29.1%	0.3%/45.1%	0.3%/90.0%	—	0.3%/34.4%

\* DBI 越小，或 DI 和 SI 越大，表示分类效果越好。最优的项目表中加粗标注。

\* 有单样例类别（MIN = 0.03%）的聚类没有意义，但评估指标会很好。表中用括号标注。

\* DBSCAN 算法在各种参数下都仅能得到一部分数据为噪声，其他全为一类的聚类结果，故未列出。

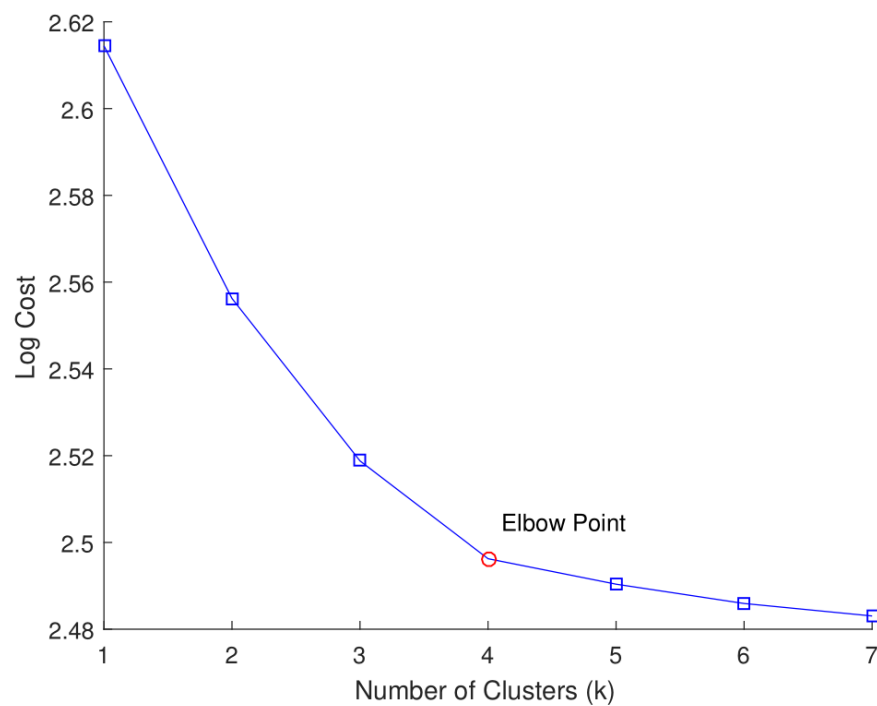
<sup>a</sup> K-medians 在算法中使用曼哈顿距离。为合理评估，指标计算中也使用曼哈顿距离。

<sup>b</sup> Mean-Shift 不接收类别数作为参数，但可以设置带宽参数得到不同的类别数的聚类。

<sup>c</sup> Affinity Propagation 算法可以自行确定类别数。实验中，给出的方案类别数为 11。

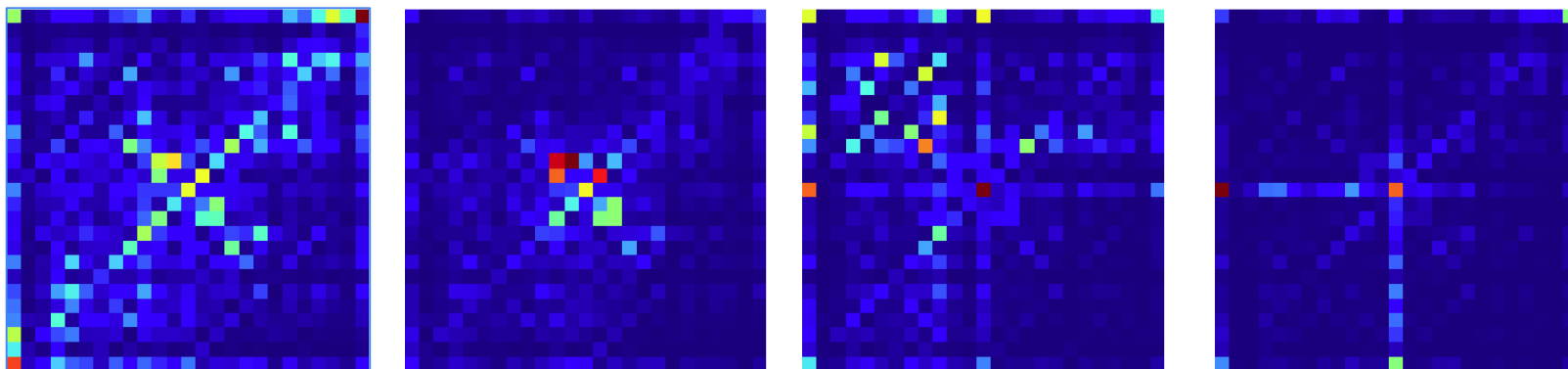
# 聚类分析 - 过程细节

- 肘点法确定类别数为4
- K-means++ 算法提供初始化聚类中心



# 结果分析 - 聚类中心

## 聚类中心可视化



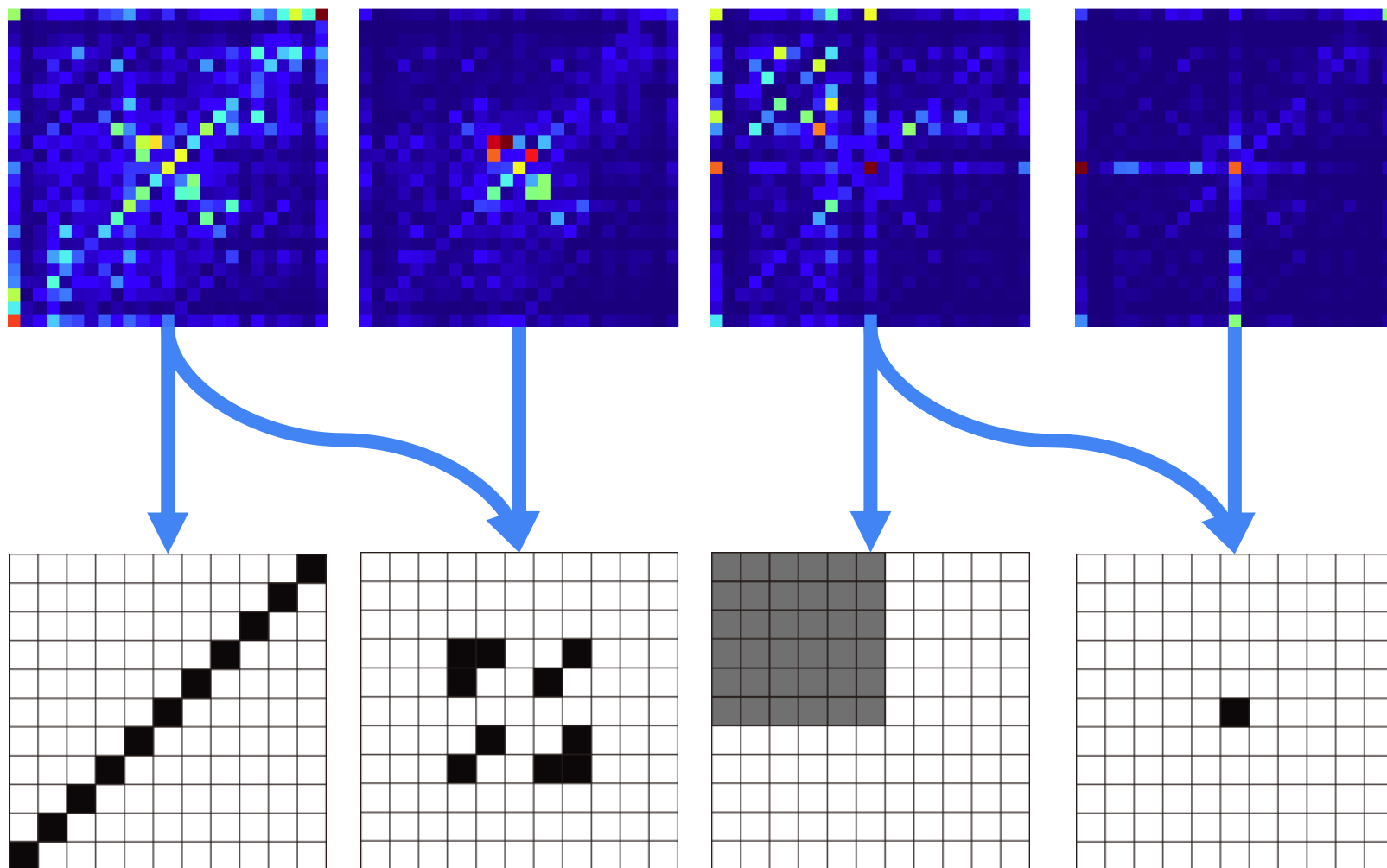
## 聚类结果指标评价

Davies-Bouldin Index	Dunn Index	Silhouette Index
2.847	0.413	0.046

评估指标较好，但缺乏相关算法对比。  
是否能够说明聚类结果足够好，有意义？

结果能否与音乐理论相对应？  
如果能，那么结果有意义。

根据Implication-Realization理论等音乐理论，可以解释如下聚类中心的风格模式



I-R 模型 registral return原则，符合听众期待的主流平和的乐曲结构。

三自然音阶连奏，旋律柔和，传达放松愉悦的心情，积极向上。

多连续下行音程，打破听众期待，宏伟并具有悲观情感。

多同音重复，具有斗争性和冲击性。

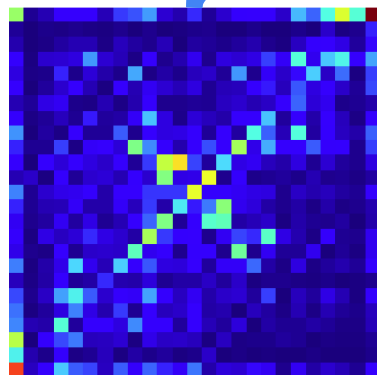
海顿



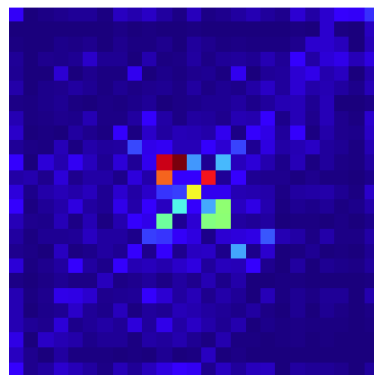
莫扎特



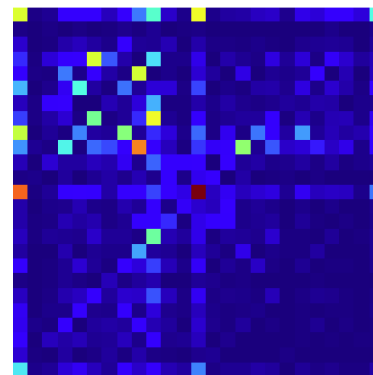
贝多芬



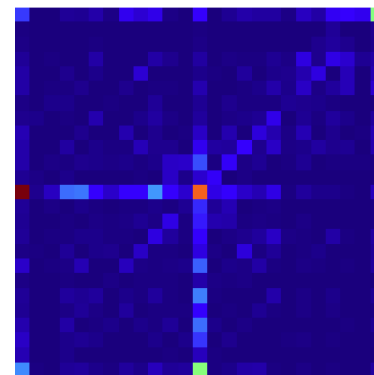
轻松愉悦的主流的混合乐曲风格



全篇柔和愉悦，积极向上的乐曲风格



悲伤、宏伟、具有冲击性的华丽作品



全篇斗争性和冲击性强烈的乐曲风格

轻松愉悦的主流曲风在三人作品中都有出现。

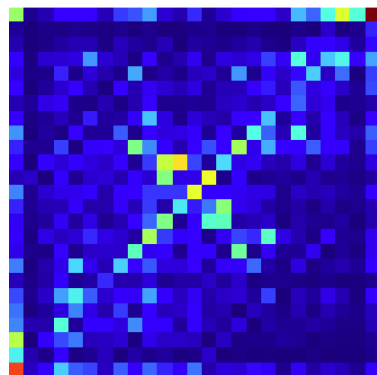
海顿



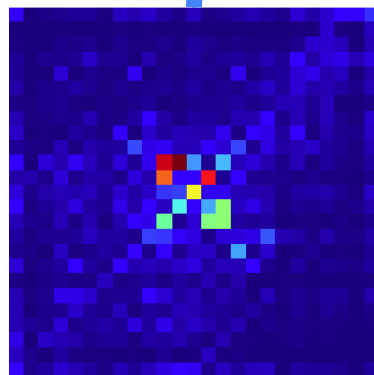
莫扎特



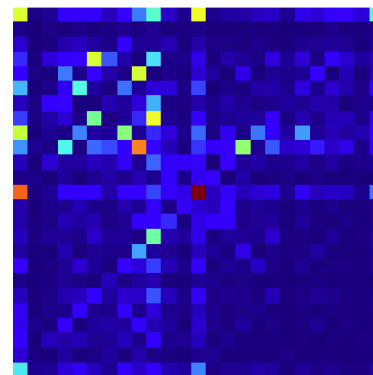
贝多芬



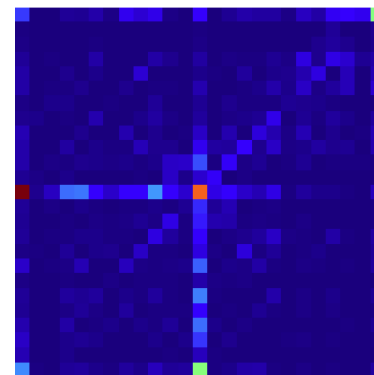
轻松愉悦的主流的混合乐曲风格



全篇柔和愉悦，积极向上的乐曲风格



悲伤、宏伟、具有冲击性的华丽作品



全篇斗争性和冲击性强烈的乐曲风格

全篇愉悦，积极向上的乐曲主要为莫扎特的作品。

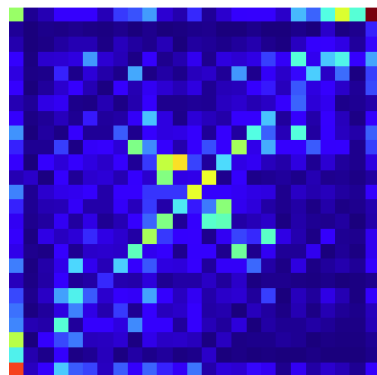
海顿



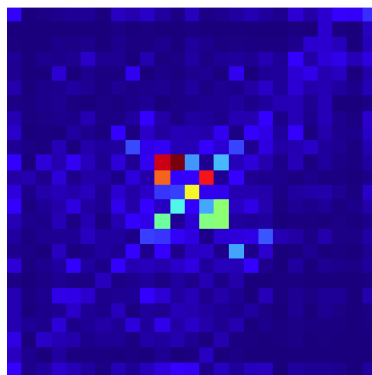
莫扎特



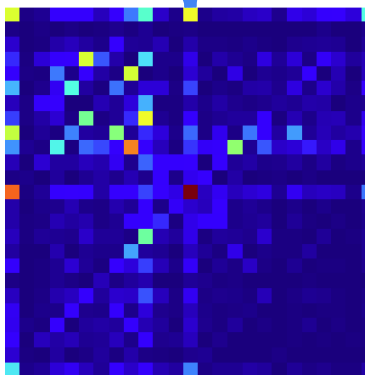
贝多芬



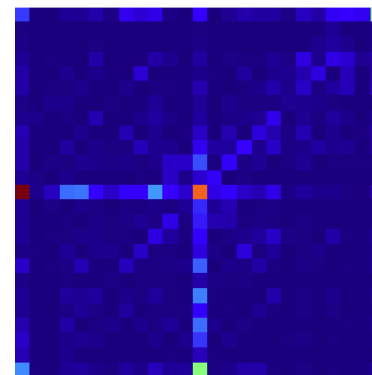
轻松愉悦的主流的混合乐曲风格



全篇柔和愉悦，积极向上的乐曲风格



悲伤、宏伟、具有冲击性的华丽作品



全篇斗争性和冲击性强烈的乐曲风格

全篇愉悦，积极向上的乐曲主要为莫扎特的作品。



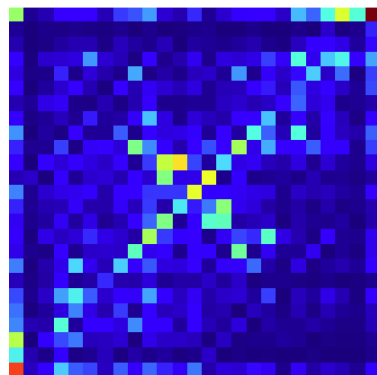
海顿



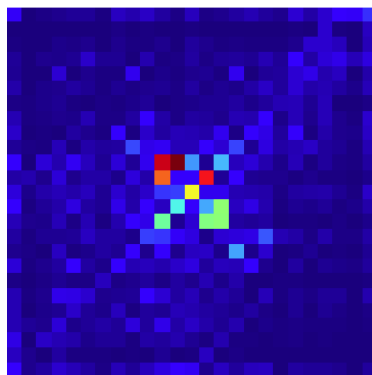
莫扎特



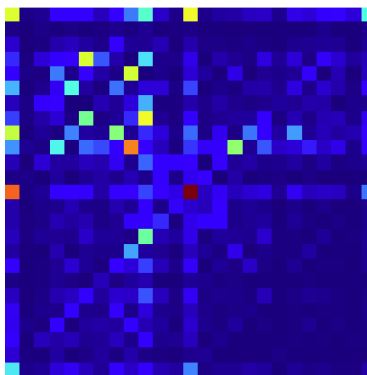
贝多芬



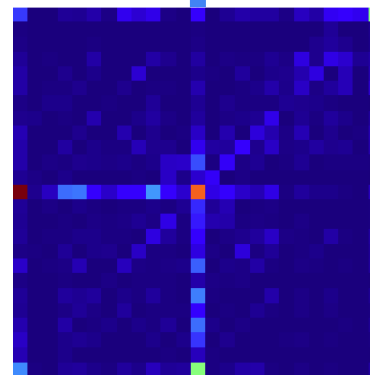
轻松愉悦的主流的混合乐曲风格



全篇柔和愉悦，积极向上的乐曲风格



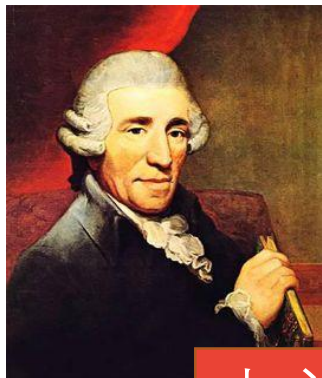
悲伤、宏伟、具有冲击性的华丽作品



全篇斗争性和冲击性强烈的乐曲风格

全篇斗争性和冲击性强烈的乐曲几乎全部属于贝多芬的作品。

海顿



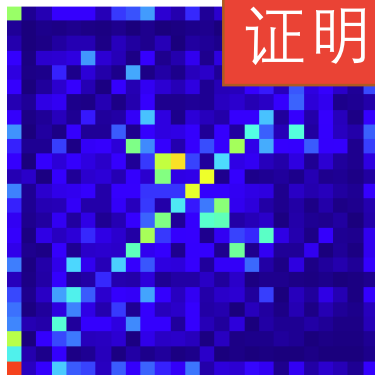
莫扎特



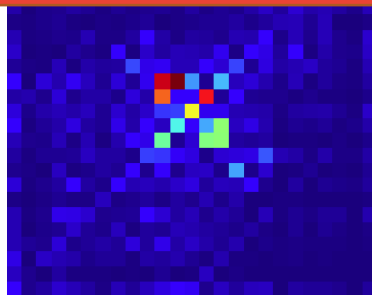
贝多芬



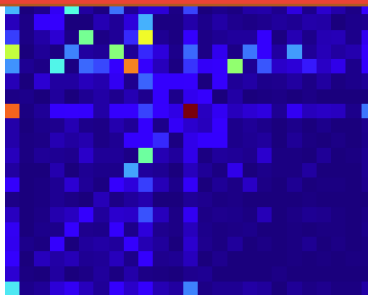
本文提出的分析方法所得结果经过了聚类指标验证，找到的特征风格与理论相符，结论与作曲家历史评价相符，可以证明方法的有效性。



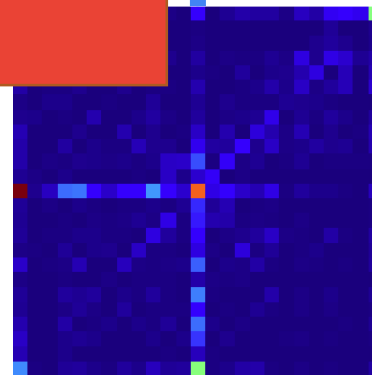
轻松愉悦的主流  
的混合乐曲风格



全篇柔和愉悦，  
积极向上的乐曲  
风格



悲伤、宏伟、具  
有冲击性的华丽  
作品



全篇斗争性和冲  
击性强烈的乐曲  
风格

全篇斗争性和冲击性强烈的乐曲几乎全部属于贝多芬的作品。

Thanks for Listening

# 工作总结

- **整理曲库数据**作为分析材料，克服已有曲库数量/格式缺陷
- 以I-R模型和相关工作作为理论依据，提出从**有结构音乐格式**中进行旋律相关的**特征提取方法**
- 实验对比了多种聚类方法，提出使用K-means算法**无监督地发现音乐风格的方法**
- 对于结果，进行了**多种指标的评价**，**结合音乐理论和音乐家的历史评价给出了相关分析**，证明和解释了方法的有效性。