

引用格式:胡嘉健,廖尚颀,陈霏.音阶法和弦表征方式对于和弦生成任务的影响[J].中国传媒大学学报(自然科学版),2023,30(03):01-07.

文章编号:1673-4793(2023)03-0001-07

# 音阶法和弦表征方式对于和弦生成任务的影响

胡嘉健,廖尚颀,陈霏\*

(南方科技大学,深圳 518000)

**摘要:**和弦是旋律伴奏的基础,为旋律搭配合适的和弦是音乐制作中不可缺少的步骤。目前深度学习已经被广泛用于和弦生成任务,但是较少研究关注不同的和弦表征方式对于和弦生成结果的影响。本文以和弦的音阶组成对和弦进行表征,将传统的和弦生成任务从单标签分类转换为多标签分类,使得和弦生成模型可以充分利用和弦的音阶信息。通过搭建LSTM网络对传统标签法与音阶法和弦生成任务进行对比,实验结果表明以音阶表征和弦的方式在和弦类别预测上优于传统的标签法,且音阶法可以提升和弦生成模型对少数类和弦的预测能力。

**关键词:**和弦表征;长短时记忆网络;和弦生成

**中图分类号:**TP18 **文献标识码:**A

## The influence of scale-based chord representation in chord generation tasks

HU Jiajian, LIAO Shangdi, CHEN Fei\*

(Southern University of Science and Technology, Shenzhen 518000, China)

**Abstract:** Chord is the basis for accompaniment to the melody, and matching the right chord to melody is an indispensable part of music making. Nowadays, deep learning has been widely used in chord generation tasks, however, few studies have studied the impact of different chord representations on the final result. In this paper, chord composition is characterized by the scale composition of the chord, and the traditional chord generation task is replaced from single label classification to multi-label classification, so that the model can make full use of the scale information of the chord. By building an LSTM network to compare the two methods, representing chords by scale is better than the traditional label method in the prediction of chord categories, and the scale method can improve the prediction ability of the model for minority chords.

**Keywords:** chord representation; LSTM; chord generation

### 1 引言

和弦是指有一定音程关系的三个或三个以上的音,按照三度或非三度关系在纵向上叠加形成的一组音。和弦是为旋律伴奏的基础,旋律与和弦相契合,

才能生成动听的音乐。和弦包括两个属性,一为其最低音的音高,称为根音,根音在时间上的变化构成了类似于旋律的线条,被称为和弦进行;另一属性为和弦类别,即和弦内部音高之间的音程关系,如大三和弦、小三和弦等。为一段旋律搭配合适的和弦,需要

**作者简介(\*为通讯作者):**胡嘉健(1998-),男,硕士研究生,主要从事计算机音乐相关研究。Email:12032208@mail.sustech.edu.cn;廖尚颀(1998-),男,硕士研究生,主要从事生理信号处理相关研究。Email:12132129@mail.sustech.edu.cn;陈霏(1975-),男,博士,教授,主要从事声学、生理信号、人机接口等相关研究。Email:fchen@sustech.edu.cn

熟悉和弦进行与和声相关的知识,选择合适的和弦根音与和弦类别,对于缺乏相关音乐知识的人很困难,因此,利用计算机自动为旋律生成和弦成为了音乐生成领域一项热点技术<sup>[1-3]</sup>。

计算机生成和弦目前存在两种方案,包括基于给定的规则或是数据集进行匹配,以及深度学习方法。第一种方案例如 Navarro-Caceres 等人提出的 ChordAIS 系统<sup>[4]</sup>,该系统通过音调距离空间<sup>[5]</sup>(Tonal Interval Space, TIS)对下一时刻的和弦类别进行预测。曹等人<sup>[6]</sup>针对蒙古族歌曲构建了具有蒙古族音乐和声特色的数据库与匹配规则,实现了为蒙古歌曲的旋律搭配和弦的目的。

基于定制规则的方案对于音乐知识储备仍有较高的要求,而深度学习的方案则通过构建模型自发从数据中学习相关乐理来达到目标。由于音乐是时序信号,最为常用的模型是长短时记忆(Long Short-Term Memory, LSTM)网络<sup>[7]</sup>。Lim 等人<sup>[8]</sup>对比了双向长短时记忆(Bi-direction Long Short-Term Memory, Bi-LSTM)网络、隐马尔可夫模型(Hidden Markov Model, HMM)以及基于深度神经网络的隐马尔可夫模型(DNN-HMM)在为旋律生成和弦任务上的差异,结果表明 Bi-LSTM 模型生成的和弦序列更加符合原音乐,该模型专注于大、小三和弦生成,不包含其余类别和弦。Leonhard 等人<sup>[9]</sup>提出了一个为爵士乐预测和弦的 LSTM 模型,该模型考虑了更多和弦类别,结果表明数据集中数量少的和弦的预测准确率也较低。

目前大部分基于深度学习模型的和弦生成任务是使用和弦标签进行和弦表征,即每一类和弦设置一个标签。这样的方式无法充分表达和弦内部音阶信息,需要模型在训练过程中进行学习。Brunner 等人<sup>[10]</sup>在研究基于和弦条件下的旋律生成任务时发现模型经过训练后,和弦标签通过 embedding 层与主成分分析(Principal Components Analysis, PCA)降维后,呈现出一个模糊的音乐中五度圈的分布规律,说明模型在训练过程中学到一定乐理知识,但其归纳能力有限。因此推测在和弦表征时按照和弦的音阶规则,以向量形式对和弦中的音阶信息进行表征,在相同训练条件下,可以提升模型的性能。Korzeniowski 等人<sup>[11]</sup>尝试了独热编码表示和弦类别以及固定长度的基于语言模型的定长向量来表示和弦类别,该工作表明了可以尝试不同方法来对最终生成的和弦进行表征,但是没有对二者具体表现进行对比。Garoufis 等人<sup>[12]</sup>在其提出的交互式和弦预测系统中,成功地以

和弦内部音阶作为目标进行吉他和弦的预测,展示了以音阶向量表示模型最终输出和弦的可行性。王静莹<sup>[13]</sup>提出的和弦旋律双编码的和弦生成模型使用四个 12 维的独热向量对和弦进行编码,同样验证了音阶向量表征和弦的可行性。

此外,大部分和弦生成任务更多是考虑和弦根音的生成,和弦类别上只有常用的大三和弦与小三和弦,合计 24 个和弦标签。而实际歌曲中还包含了很多出现频次较少的和弦。使用向量形式进行和弦表征,可以以定长维度的向量表示尽可能多的和弦类别。

为了探究以音阶向量进行和弦表征的方法与传统标签法在和弦生成任务中的性能差异,本文搭建了为流行音乐旋律生成和弦的 LSTM 模型,该模型以和弦音阶作为生成目标,将和弦生成任务从单标签多分类转换为了多标签二分类,并与传统的单标签多分类方法对比,实验结果发现该方法在预测少数和弦以及和弦类别分类上有更好的效果。

## 2 方法

### 2.1 和弦表征方法

为了探究以向量表征和弦的方式对和弦生成任务的影响,实验中对比设置了标签法与音阶法进行表征和弦。标签法与音阶法的和弦表征方法如下。

**标签法:**即用一个  $12 \times N$  维独热向量表示,  $N$  是和弦的种类数, 12 表示根音数目。如图 1 展示了在仅包含大三和弦与小三和弦类别的数据集中 C:maj 和弦的表示方法。最终每段音乐和弦表示为一个  $(Len, 12 \times N)$  的矩阵,  $Len$  表示乐句长度。

**音阶法:**用一个 12 维的独热向量和一个 12 维的多热向量表示和弦。和弦的类别是由其内部组成音阶决定的,在一个八度内共有 12 个音高,所以可以通过独热向量表示和弦根音属性,多热向量表示和弦类别。例如 C:maj 和弦是大三和弦,根音 C 的表示方法为:  $[1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]$ , 大三和弦三个音的音程关系为大三度和小三度,即在 0、4、7 三个位置为 1,其余为 0,多热向量表示为  $[1\ 0\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ 0]$ ,将二者拼接,如图 2 所示。最终每段音乐的和弦表示为一个  $(Len, 24)$  的矩阵。

音阶法是将和弦的根音与类别分开进行记载,标签法则是二者不区分记载。为了控制变量,新设定了双标签法。

**双标签法:**采用了两个独热向量的方式,分别记录和弦的根音与和弦类别。假设在数据集中和弦类别共有6类,即这两个独热向量维度分别为12和6,则向量如图3所示。最终和弦表示为(Len, 12+6)的矩阵。

Maj												Min													
C	#C	D	#D	E	#E	F	#F	G	#G	A	#A	B	C	#C	D	#D	E	#E	F	#F	G	#G	A	#A	B
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

图1 C:maj和弦的标签法示例

C	#C	D	#D	E	#E	F	#F	G	#G	A	#A	B	C	#C	D	#D	E	#E	F	#F	G	#G	A	#A	B
1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	

图2 C:maj和弦的音阶法示例

C	#C	D	#D	E	#E	F	#F	G	#G	A	#A	B	Maj	Min	dim	maj7	min7	7
1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0

图3 C:maj和弦双标签法示例

## 2.2 数据集与预处理

本文选取了POP909数据集<sup>[14]</sup>,该数据集包含了近60年来中国流行歌曲的MIDI文件909份,并且让专业音乐家重新进行了钢琴和弦伴奏的编排。每个MIDI文件包括了人声旋律,主奏乐器旋律和钢琴和弦伴奏三个轨道,还额外以文本形式标注了调式、和弦信息与节拍信息,这些信息在时间上都与原始音频对齐。本文提取了人声旋律轨道中的信息,并转换为pianoroll数字矩阵格式,和弦信息则从标注的文本文件中提取。

与传统的标签记录和弦的方法相比,音阶表示法可以以较少的维度表示几乎所有的和弦。为了充分探究两种方法在和弦生成任务上的差异,本论文参考了Carsault等人<sup>[15]</sup>的工作,制作了三组包含不同和弦种类数目的数据集A0、A1、A2,每组包含的和弦种类如表1所示。A0只包含大、小三和弦两种,A1有合计6种和弦,A2包含14种和弦。在和弦生成任务中,难度也是依次上升的。每个数据集按照8:1:1的比例划分为了训练集、验证集与测试集,并进行了交叉验证实验。

表1 数据组和弦类别详情

数据组	含有的和弦类别
A0	maj, min
A1	maj, min, dim, maj7, min7, 7
A2	maj, min, dim, aug, maj6, min6, maj7, minmaj7, min7, 7, dim7, hdim7, sus2, sus4

对于数据集中每一首歌,首先通过其每分钟节拍

数(Beat Per Minute, BPM)将MIDI文件中的人声轨道转换为单位时长为十六分音符的pianoroll格式的数字矩阵。随后以16拍为长度,8拍为步长进行平移裁剪乐句,若裁剪的乐句中旋律发声时长小于8拍则舍弃该乐句。即最终乐句是用[64,128]的0,1数字矩阵表示,64是时长,128是音高区间。由于旋律的音高分布并没有布满整个音高区间,在最终旋律表征上,则使用了一个长度64的数值向量,数值范围为0-49,0-47表示了一个范围为48音高,48表示音符的持续,49表示休止音符。另外,参考Choi等人的工作<sup>[16]</sup>,节奏信息的增加可以提高模型最后生成结果中旋律与和弦的适配程度,所以再额外增加了旋律的韵律信息输入,它记录着每个音符开始的位置。这是一个长度为64的向量,每一个向量位置数值范围为{0, 1, 2},0表示休止,1表示音符的保持,2表示音符开始。

POP909数据集中的和弦标注十分精确,某些和弦会增加额外的音符或者低音,额外的音符会被记载在括号中,低音则记载在斜杠后,如果将这些全部记录为新的和弦,种类会过于繁多,且数量分布十分稀疏。因此这些携带额外音符或是低音的和弦类别被归为了其原本的和弦类别。比如F:maj7(11)/3表示了包含11音以及3音低音的F:maj7和弦,这一类和弦标签会被归类于其本体F:maj7和弦。经过预处理后收集到的和弦分布如图4所示。

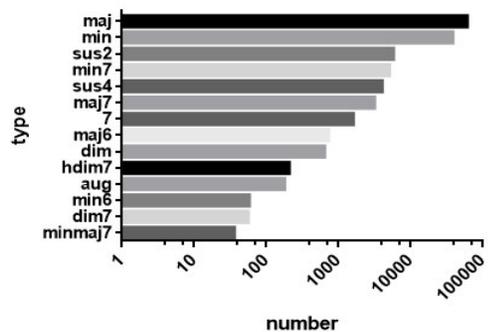


图4 POP909数据集中各和弦类别数目

## 2.3 模型与损失函数的选取

本研究参考了Lim等人<sup>[8]</sup>工作,选用了Bi-LSTM网络搭建模型。音高和韵律信息都是独热向量,先通过embedding层进行编码,再输入到LSTM中。LSTM输出连接一个全连接层的解码器,通过解码器获得最后和弦种类。在标签法中,最后输出层为(64, 12×N)的矩阵,64为时间长度,N为和弦种类,标签法

因为是单标签的分类,损失函数采用了交叉熵损失函数(Cross Entropy Loss)。音阶法中,最后输出的是(64, 24)的矩阵,这是一个多标签的分类,每个标签都有可能是0或1,所以采用了二分类交叉熵损失函数(Binary Cross Entropy Loss)。双标签法最终输出是(64, 12+N)的矩阵,和标签法一样采用了交叉熵损失函数。A2数据组中和弦类别有14类,分布不均衡现象对少数类结果影响大,因此在这一组实验中还尝试了Focal Loss损失函数对类别不均衡问题进行优化。普通交叉熵损失函数如公式(1)所示:

$$Loss = L(y, p) = -y \log(p) - (1 - y) \log(1 - p) \quad (1)$$

式中 $y$ 为标签,在二分类中对应0与1; $p$ 为模型预测样本为正的的概率,范围在0-1。将 $p_t$ 定义为:

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases} \quad (2)$$

则交叉熵损失函数如下:

$$Loss = -\log(p_t) \quad (3)$$

实际的Focal Loss是交叉熵损失函数的以下变体:

$$FL(p_t) = -\alpha(1 - p_t)^\gamma \log(p_t) \quad (4)$$

其中 $\alpha$ 是类别权重因子,对于占比更大的类别,使其对最终的损失函数贡献小,让模型更关注那些类别稀少的样本,从而达到均衡的目的。 $(1 - p_t)^\gamma$ 中 $\gamma$ 为难度权重因子,可以根据区分该样本的难度给予不同的权重,当 $p_t$ 趋向于1的时候,说明模型对该样本区分为正样本或是负样本有着较大把握,所以最终会给出一个较小的权重,表示这个样本区分难度较小,可以减少其对最后损失函数的贡献。以此途径,Focal Loss还有对难易程度不同的样本区分训练的能力。通过参考Focal Loss原文中权重设置与进行的部分对照实验,本实验最后设置 $\gamma = 2$ ,标签法中出现频率高的maj与min和弦设置权重 $\alpha$ 为0.2,其余为0.8。音阶法中则是根据每个音阶出现0、1的差异频次为12维标签分别设置权重,12个音类别均值如表2所示,最终权重设置为[0.2, 0.2, 0.8, 0.5, 0.5, 0.8, 0.8, 0.2, 0.8, 0.8, 0.8, 0.8]。

表2 音阶法各音阶0、1类别均值

C	#C	D	#D	E	F
1	0	0.048	0.368	0.550	0.023
#F	G	#G	A	#A	B
0.017	0.991	0.001	0.007	0.057	0.037

模型框架示意图如图5所示,模型与实验参数设置如表3所示。训练使用Adam优化器进行网络参数更新。

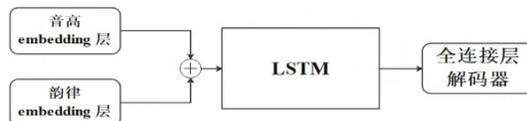


图5 实验模型结构示意图

表3 模型与实验参数设置

模型参数	音高 embedding 层维度	128
	韵律 embedding 层维度	128
	LSTM 隐藏层维度	256
	LSTM 层数	1
实验参数	最大训练轮次	250
	批量大小	64
	学习率	0.0005

### 3 结果与分析

#### 3.1 评估方式

本课题通过比较不同方法训练出的模型在测试集中和弦预测的准确率与F1分数来定量分析。准确率是预测和弦与真实和弦之间匹配的样本数量与样本总数的比值。由于模型最终输出是与输入乐段等长的一段矩阵或向量,即每一单位时间(十六分音符长度)都会有一个和弦类别结果,在最终的音乐生成时,和弦是以小节为单位进行生成的,每小节的最终和弦是该小节出现次数最多的和弦类别。为了更好地衡量不同模型之间性能的差异,本文将每一单位时间的结果皆纳入统计。

另外为了比较模型对少数类和弦分类的效果,还使用了F1分数来衡量模型精确度。F1分数是通过准确率(Precision)与召回率(Recall)计算的,TP表示预测为1,实际为1的样本数量;FP表示预测为1,实际为0的样本数量;FN表示预测为0,实际为1的样本数量;TN表示预测为0,实际为0的样本数量。准确率公式如公式(5)所示,召回率公式如公式(6)所示,F1分数计算公式为公式(7):

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (7)$$

对于多分类任务,需要计算每一个类别的F1分数。本文选择了macro-F1分数作为评判标准,macro-

F1分数是通过计算每一类别的F1值并取平均后得到,macro-F1分数不受样本种类数目不平衡的影响,更符合该任务的评价指标。

### 3.2 实验结果

各组训练情况相似,损失函数数值皆在200轮训练后趋于收敛,各组统一选取200轮训练结果进行对比。

#### (1)A0数据组实验结果

由于在A0数据集中,和弦类别仅有大三和弦与小三和弦,且大三和弦与小三和弦在音阶表示上仅仅相差一个单位,双标签法在A0数据集上与音阶法无理论上差异,所以没有进行双标签法在A0数据集上的实验。且在A0数据组中,没有和弦类别不均衡的情况,因此只展现了准确率信息。表4展示了音阶法、标签法的精度结果,总准确率表示和弦根音与类别皆预测准确。

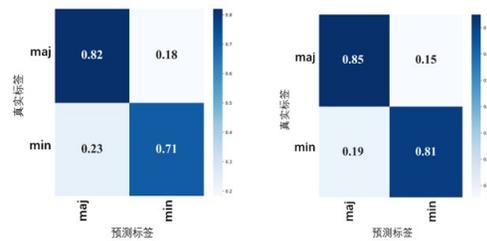
由表4可知音阶法的根音准确率71.12%优于标签法的70.27%,类别准确率82.30%优于标签法的78.94%,而整合后得准确率为65.3%,略低于标签法。在标签法中,训练中直接优化目标的是和弦的总标签损失函数,它包括根音与类别两部分,但是其内部并不会考虑和弦的内部构成音阶因素与根音因素,所以其单独的根音与类别准确率不如音阶表示法。而音阶法后12维表示和弦类别的向量中,大三和弦与小三和弦的区别也仅在3、4维位置上,在A0数据组上,

二者结果相差不大。

表4 A0数据组实验结果统计(%)

	总准确率	根音准确率	类别准确率
标签法	66.67	70.27	78.94
音阶法	65.30	71.12	82.30

图6展示了音阶法与标签法在类别分类上的混淆矩阵,从图中可以看出,音阶法的大三和弦与小三和弦识别率均高于标签法。



(a) 标签法

(b) 音阶法

图6 A0数据组实验类别准确率混淆矩阵

#### (2)A1数据组实验结果

从A1数据组开始,双标签法加入实验并进行对比,这一组数据中,和弦总共有maj、min、dim、maj7、min7和7和弦六种。表5展示了三种方法预测准确率以及相应的macro F1分数。从整体准确率上看,标签法的总准确率为65.21%,高于另外两种方法,但其F1分数低于双标签法。根音准确率则是双标签法最高,为69.40%,且其F1分数也最高。类别准确率上则是音阶法最高,为78.43%,而F1分数是双标签法最高。

表5 A1数据组实验结果统计

	总准确率(%)	总 macro F1	根音准确率(%)	根音 macro F1	类别准确率(%)	类别 macro F1
标签法	65.21	0.4348	69.31	0.6932	76.37	0.4844
双标签法	62.36	0.4394	69.40	0.6941	78.14	0.5407
音阶法	61.95	0.4103	68.56	0.6902	78.43	0.5112

图7展示了三种方法在和弦类别预测上的混淆矩阵。其中maj与min和弦的准确率最高,且其他类的和弦也大多被预测为这两类。但模型对少数类和弦仍有部分预测能力。图中音阶法多出的维度6代表被分类到该数据组中不存在的和弦,结果显示音阶法中,被错误分类到不存在的和弦类别比例小,证明音阶法模型可以综合和弦的音阶信息从而对和弦类别进行分类。

#### (3)A2数据组实验结果

A2数据组涵盖了POP909中几乎全部的和弦种类,

类别不均衡情况严重,所以在标签法与音阶法上增加了Focal Loss损失函数均衡类别间数目的不平衡。表6展示了三种方法以及增加了Focal Loss后的音阶法与标签法,合计五种方法在A2数据组上的最终结果。总准确率仍然是标签法最高,为61.15%,但是音阶法的总准确率也十分接近,为61.08%。且音阶法在其余各项指标中均高于标签法,说明音阶法对少数类和弦有更好的预测能力。同样,双标签法也通过将和弦根音与类别分开,获得了比标签法更好的对少数和弦类别的预测能力,且类别F1分数略高于音阶法,其余低于音阶法。

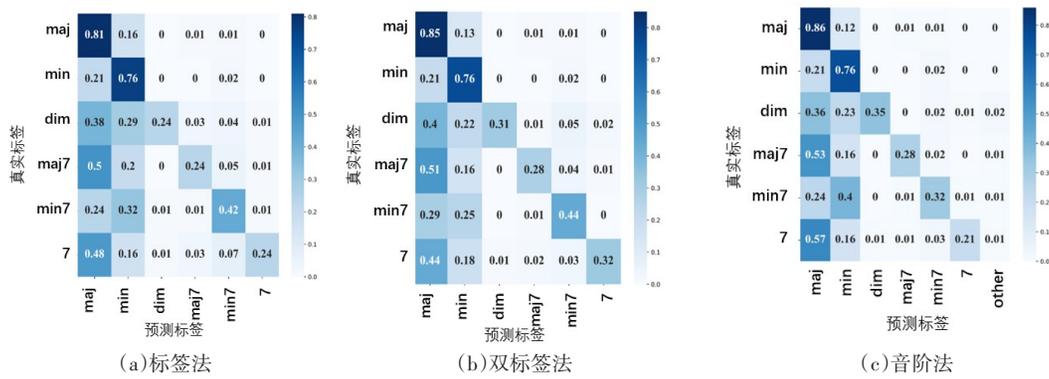


图7 A1数据组实验类别准确率的混淆矩阵

表6 A2数据组实验结果统计

	总准确率(%)	总macro F1	根音准确率(%)	根音macro F1	类别准确率(%)	类别macro F1
标签法	61.15	0.2745	69.50	0.6946	70.70	0.3582
双标签法	60.21	0.2919	70.79	0.7079	73.51	0.4473
音阶法	61.08	0.3092	71.65	0.7199	74.19	0.4311
标签法+Focal Loss	55.06	0.3291	63.92	0.6387	66.50	0.4349
音阶法+Focal Loss	58.77	0.2242	72.66	0.7288	70.47	0.3125

另一方面,增加了Focal Loss后,标签法的总准确率下降,类别准确率下降,相应的总F1分数与类别F1分数提高,根音准确率与F1分数下降。Focal Loss牺牲了模型对多数和弦(大三和弦与小三和弦)的预测能力,提升了对少数和弦的预测能力。而在音阶法中,总准确率、类别准确率,以及其相应的F1分数都

下降,但是根音的准确率与F1分数上升,说明Focal Loss在该权重系数上无法优化模型对少数和弦的预测能力,反而会产生反效果,同时导致了模型更加关注根音上的损失函数,使得根音的预测结果提升。

图8展示了三种基础方法的混淆矩阵。音阶法混淆矩阵图中的第14维表示未在数据集中的错误和弦类

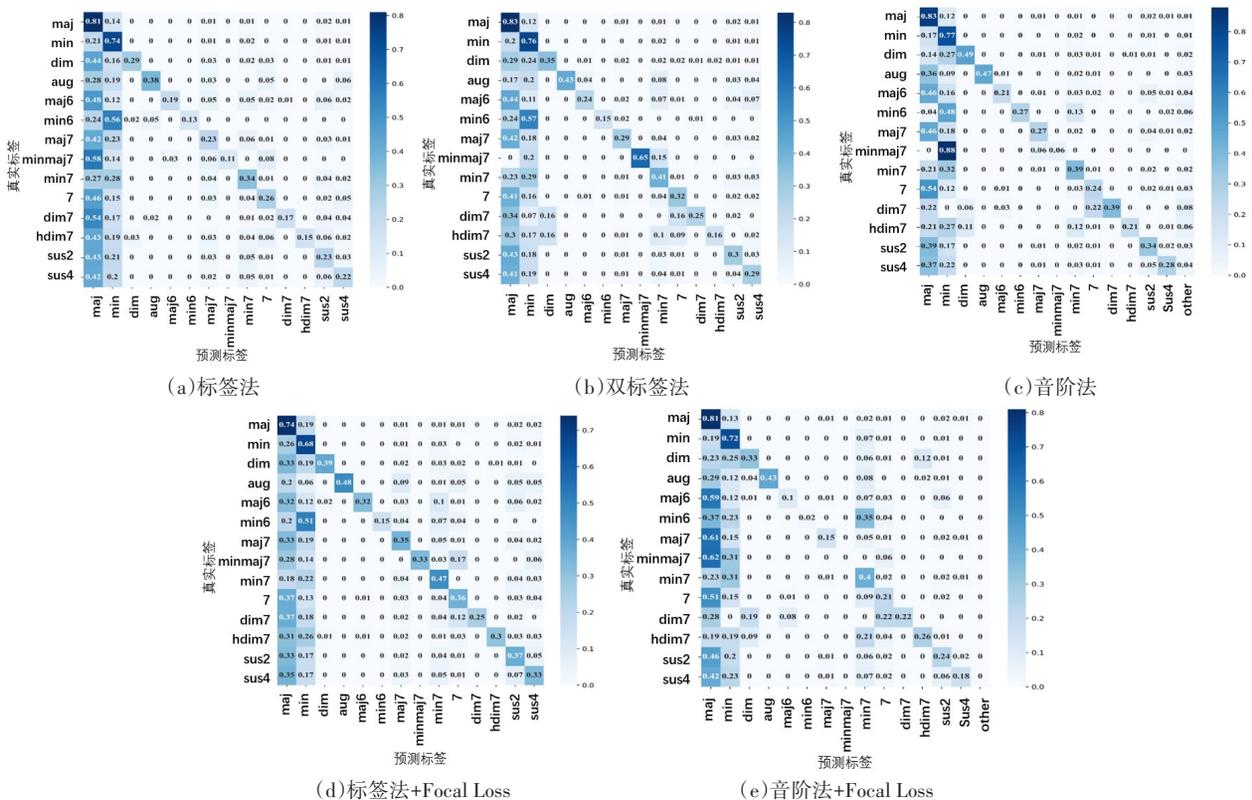


图8 A2数据组实验类别准确率的混淆矩阵

别。从图中可见,在主对角线数值上,音阶法、双标签法以及增加了Focal Loss的标签法普遍高于标签法,而增加了Focal Loss的音阶法的主对角线数值普遍有所降低。

## 4 结论

本文对比了标签法与音阶法在和弦类别数目不同的数据集下对旋律预测和弦的准确率。在总准确率上,标签法效果更好,但是在A1数据组中,双标签法有着比标签法更高的和弦类别与根音预测准确率以及F1分数,说明将和弦类别与根音拆分进行预测可以一定程度缓解数据集中的和弦不均衡导致的模型对少数类和弦识别能力差的问题。

在和弦类别最多的A2数据组,标签法与音阶法的总准确率相差不大,且音阶法的根音准确率、类别准确率以及相应的F1分数皆高于标签法,双标签法也仅有类别F1分数略高于音阶法,说明以音阶信息取代标签可以提升模型性能。

Focal Loss对于标签法有着较大的提升,可以使模型更加关注稀少类别的和弦。另外Focal Loss不适用音阶法,音阶法本身将和弦类别拆分成12维的多标签的二分类任务,在一定程度上已经做了平衡样本的工作。虽然这样的多标签依旧存在着二分类不平衡的状态,但是这些标签相互之间不是独立的,是具有一定联系的,所以盲目的设定权重与调节系数只会得到负面效果。

实验结果显示使用12维的音阶来替代传统的和弦类别标签作为模型最终输出的音阶法,可以利用和弦的音阶信息提升模型对和弦类别的预测能力,且当模型需要预测的和弦类别增多时,比标签法更能适应数据集中和弦类别不均衡的情况。

本文在和弦表征上没有探究单独为和弦训练一组用于表征的编码器解码器对生成任务的影响,在后续工作中,可以进一步考虑该方法,将和弦的类别映射到连续的高维特征空间,并使用数据增强的方法来均衡样本,从而优化最终结果。

## 参考文献(References):

- [1] Briot J P, Hadjeres G, Pachet F D. Deep learning techniques for music generation[M]. Berlin: Springer, 2020: 1-10.
- [2] Hernandez-Olivan C, Beltran J R. Music composition with deep learning: a review[J]. Advances in Speech and Music Technology: Computational Aspects and Applications, 2022: 25-50.
- [3] 陈吉尚, 哈里旦木·阿布都克里木, 梁蕴泽, 等. 深度学习在符号音乐生成中的应用研究综述[J/OL]. 计算机工程与应用, 2023, 59(9): 27-45.
- [4] Navarro-Caceres M, Caetano M, Bernardes G, et al. ChordAIS: an assistive system for the generation of chord progressions with an artificial immune system[J]. Swarm and Evolutionary Computation, 2019, 50: 100543.
- [5] Bernardes G, Cocharro D, Caetano M, et al. A multi-level tonal interval space for modelling pitch relatedness and musical consonance[J]. Journal of New Music Research, 2016, 45(4): 281-294.
- [6] 曹西征, 牛靖雯, 吴瑞琪, 等. 面向蒙古族歌曲旋律的自动伴奏算法[J]. 计算机工程, 2016, 42(07): 304-309.
- [7] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [8] Lim H, Rhyu S, Lee K. Chord generation from symbolic melody using BLSTM networks[C]//Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR), 2017: 621-627.
- [9] Driever L, Harasim D, Jagt M L, et al. Improving chord prediction in Jazz music using melody information[C]//Proceedings of the 19th Sound and Music Computing Conference, 2022: 147-154.
- [10] Brunner G, Wang Y, Wattenhofer R, et al. JamBot: music theory aware chord based generation of polyphonic music with LSTMs[C]//Proceedings of the 29th International Conference on Tools with Artificial Intelligence (ICTAI), 2017: 519-526.
- [11] Korzeniowski F, Sears D R W, Widmer G. A large-scale study of language models for chord prediction[C]//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018: 91-95.
- [12] Garoufis C, Zlatintsi A, Maragos P. An LSTM-based dynamic chord progression generation system for interactive music performance[C]//Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020: 4502-4506.
- [13] 王静莹. 基于和弦与旋律双编码的音乐生成算法[J]. 信息与电脑(理论版), 2022, 34(07): 98-100.
- [14] Wang Z, Chen K, Jiang J, et al. Pop909: a pop-song dataset for music arrangement generation[C]//Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR), 2020: 38-45.
- [15] Carsault T, Nika J, Esling P. Using musical relationships between chord labels in automatic chord extraction tasks[C]//Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR), 2018: 18-25.
- [16] Choi K, Park J, Heo W, et al. Chord conditioned melody generation with transformer based decoders[J]. IEEE Access, 2021, 9: 42071-42080.