DOI:10.16451/j.cnki.issn1003-6059.2013.07.002

一种基于液体状态机的音乐和弦序列识别方法*

张冠元[1,2] 王 斌[1]

[1](中国科学院计算技术研究所 北京 100190)
[2](中国科学院大学 北京 100049)

摘 要 文中提出一种基于液体状态机的音乐和弦序列识别方法. 该方法首先将音乐信号进行切分采样并对每帧提取音级轮廓(PCP),经训练后得到一个液体状态机模型. 方法提出两类奇异矩阵、和弦出现概率向量、和弦变换矩阵,它们可用在和弦序列后处理阶段. 在神经网络模型、隐马尔科夫模型、回声状态网络模型、液体状态机模型上进行的初步实验得到 8 组实验数据. 数据表明液体状态机模型对音乐和弦序列具有较好的识别效果,文中提出的后处理算法也能显著提高识别准确率.

关键词 和弦序列识别,液体状态机,音乐信息检索,模式分类
中图法分类号 TP 391.4

Liquid State Machine Based Music Chord Sequence Recognition Algorithm

ZHANG Guan-Yuan[1,2], WANG Bin[1]
[1](*Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190)
[2](*University of Chinese Academy of Sciences, Beijing* 100049)

ABSTRACT

A chord sequence recognition algorithm based on Liquid State Machine (LSM) is presented. Firstly, the music signal is segmented and Pitch Class Profile feature is extracted for every frame. Then, a LSM model is achieved after training. Two kinds of Bizarre Chord, chord appears probability vector and chord transformation matrix, are presented to post-process the chord sequence outputted by LSM. 8 sets of experimental data from neural network model, hidden Markov mode, echo state network model and LSM model show that the LSM gets a good performance, and the post-processing method also effectively improves the recognition accuracy.

Key Words Chord Sequence Recognition, Liquid State Machine, Music Information Retrieval, Pattern Classification

* 国家自然科学基金项目(No. 61070111)、中国科学院先导项目课题(No. XDA06030200)资助
收稿日期:2013-02-04;修回日期:2013-04-16
作者简介 张冠元(通讯作者),男,1986 年生,博士研究生,主要研究方向为信息检索. E-mail: zhangguanyuan@ict.ac.cn. 王斌,男,1972 年生,副研究员,博士生导师,主要研究方向为信息检索.

1 引 言

和弦是音乐中具有一定音程关系的一组音符. 按照三度叠置关系,将 3 个或更多的音符纵向结合就形成和弦. 和弦能丰富音乐的旋律表达、体现音乐的情感变化. 音乐信息检索领域中的很多问题,如音乐相似度匹配、音乐流派分类、音乐情感分析,都需分析音乐的谐波结构. 和弦是体现音乐谐波结构的基本单位.

和弦识别是根据给定的音乐,识别出音乐中每个音节的和弦,最后输出一个和弦时间序列. 目前和弦识别面临的主要问题包括: 1) 大部分的音乐都不是单乐器乐曲,从混叠发声的多声部音乐中识别每个音高,准确分析复杂的谐波结构难度较大; 2) 从生物学角度来看,人类听觉神经系统对于音乐信息的处理过程及对音乐属性的分辨过程还是未知的.

为解决这些问题,研究者们尝试使用多种模型. 文献[1]、[2] 以音级轮廓(Pitch Class Profile, PCP) 作为音频特征,训练隐马尔科夫模型(HMM) 进行分类,对音乐中的和弦序列进行识别. 文献[3] 使用高斯模型,并使用一些启发式规则来提高效果. 一般来说,每个节拍的和弦是不同的,该文利用这一规律,在节拍的间隙进行和弦识别,取得较好效果. 文献[4] 基于主旋律往往故意包含非谐波音符的假设,通过添加一个衰减乐曲主旋律的预处理步骤来提高识别效果. 文献[5] 在 12 维音级轮廓的基础上提出 6 维音色重心向量,并训练得到一个 HMM,取得较好效果. 文献[6] 提出音乐中的鼓声等敲击声是非谐波的,它对和弦检测造成干扰,因此添加一个预处理步骤用于去除音乐中的敲击声音.

文献[7]、[8] 根据和弦的基音区分方法,定义24 个基音,为每个基音建立一个 HMM,构造 Key-Dependent HMM 模型. 对于输入的音乐信号,系统利用 Viterbi 解码在 24 个模型中选择一个最大可能的基调模型,从而确定输入音乐的基调,并从对应于这个基调模型的最优状态路径来得到和弦序列.

文献[9]、[10] 均使用人工神经网络(Artificial Neural Networks, ANN) 的方法解决和弦识别问题. 文献[9] 分析音乐认知心理学,基于 ANN 构建一种和弦识别方法. 该方法使用时频变换的方式处理音乐信号,并定义一种称为音级分布矩阵的新特征,通过半监督学习方法识别和弦. 在自己构建的数据集上进行实验,得到 57% 的识别准确率.

本文在和弦识别分类算法方面开展研究,针对和弦识别研究中常用的 HMM、高斯模型以及 ANN 存在的不足,创新性地选用了液体状态机作为模式分类模型. 在音频特征方面,本文选用音级轮廓作为音频特征,并提出一种和弦序列后处理方法,对模型的输出序列进行优化,最后通过实验对比分析 HMM 方法、传统神经网络方法、回声状态网络方法(Echo State Network, ESN) 和本文方法的识别结果.

2 基于液体状态机的和弦序列识别算法

2.1 分段采样

在进行和弦识别之前,先将乐曲切成固定时间长度的帧,然后对每帧进行识别. 窗口大小的设定对于和弦识别准确率有直接影响. 如果窗口的长度小于一个和弦的演奏时间长度,则识别准确率将受影响,因此,较长的窗口更有利于和弦识别. 经过对 400 首音乐的统计得到以下规律:

1) 和弦切换是在节拍之间或半拍时进行的;

2) 音乐的节奏一般保持在每分钟80~120 个节拍;

3) 相邻两个和弦起始音符的间隔在 600ms ~ 750ms 之间;

4) 对于两个相邻的和弦,后一个和弦起始音符的能量明显大于前一个和弦结束音符的能量.

较长的帧长度可提高和弦识别率,但也带来另外的问题: 如果一个和弦帧边界附近发生变化,和弦就会被正确检测到,因为帧内大部分是单一的和弦; 如果和弦的变化发生在帧的中间,那么识别结果误差就会很大,因此相邻两个帧之间有一定的重叠. 本方法的帧窗口示意图如图 1 所示. 图中帧窗口长度为 $l = 600\text{ms}$,邻帧窗口重叠长度为 $l - h = 400\text{ms}$.

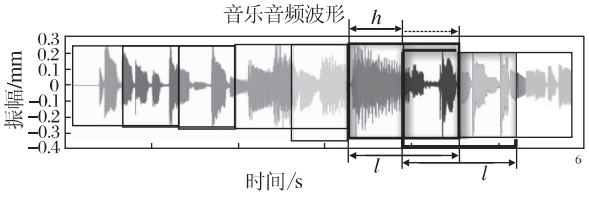

图 1 帧窗口划分示意图

Fig. 1 Diagram of frame window segmentation

2.2 PCP 特征

色度向量(Chroma Vector) 也被称为音级轮廓(Pitch Class Profile),是 Fujishima[11] 在 1999 年提出的用于和弦识别系统的特征. PCP 基于音乐的十二音律理论,是音乐信号在时频变换后的对数表示. 在

十二音律理论中,一个八度音程中有 12 个音级(C, C#,D, D#,E,F,F#,G,G#,A,A#,B) . PCP 提取短时傅里叶变换频域中的属于同一个基频类的部分,与 12 个音级相对应,形成 12 维向量,每维向量代表该音级中的累积振幅.

2.3　识别方法

2002 年,Mass 等[11] 为解决时序信号的实时计算问题提出液体状态机(Liquid State Machines, LSM) . 目前,LSM 在多个领域都取得成功应用,如时间序列预测、模式分类、语音识别、事件检测、图像处理、飞行器控制等.

LSM 的结构类似于平常的神经网络,可分为输入层、输出层和中间映射层. 中间映射层也叫液体池,是由数量庞大的神经元排列成点阵结构的回归模型,是一种神经元回路结构. 神经元之间的连接权重是在构建液体池时随机确定的,只有少量的神经元连接权重是可变化的,大部分权重是常量,固定不变,这使得权重的更新缓慢,被称为该模型的液体性质. 液体性是液体状态机与传统神经网络最大的不同. 液体状态机神经元回路是仿照大脑皮层中的回路设计的,相关实验证明,这种神经元回路在处理时序信号等方面能较好模拟人脑,且具有强大的计算能力.

液体池一般是一个长方体结构,各个神经元分布其中. 神经元之间的连接公式定义如下:

$$P(i,j) = C^* \exp\left(-\frac{D(i,j)}{\lambda^2}\right),$$

其中,$P(i,j)$ 表示神经元 i 与神经元 j 之间发生连接的概率,$D(i,j)$ 表示长方体的液体池中两个神经元的欧式距离. 参数 C 用来控制液体状态机模型的类型,偏兴奋型或抑制型,实际上表示神经元连接疏密程度.

LSM 的输入层到液体池连接是随机确定的,一个输入神经元可跟任意数目的中间层神经元相连接. 液体池中的神经元回路结构将输入序列 $u(\cdot)$ 映射到高维空间 $x^L(t)$:

$$x^L(t) = (G^L u)(t) .$$

与支持向量机(Support Vector Machine, SVM) 原理一致. G^L 可看作一种滤波机制,将输入信号投射到高维空间,使得数据更易分类. 由于液体池是一种回归神经网络,具有记忆性,$x^L(t)$ 不仅是对当前输入信号的映射,还反映历史记忆信息,因此 LSM 更适合处理动态问题. LSM 的输出层将中间状态 $x^L(t)$ 变为输出序列 $y(t)$.

输入序列从 LSM 输入层输入,可看作持续的输入流注入到液体层,在任何时刻都可读取液体状态,并通过输出函数将目标输出与液体状态对应. LSM 被看做大量的复杂有限状态机的组合,其内部状态不需定义,通过训练输出单元来改变液体层状态.

ANN 是生物神经系统的工程模拟,很多研究把它用作模拟人脑神经系统处理和弦感知信息的过程. 但传统神经网络一般使用梯度下降方法学习,而循环神经网络复杂度越高越容易造成梯度下降法的反向误差传播失效,因此传统神经网络的规模一般较小. LSM 液体池的构成与传统神经网络有较大差别,是大规模的随机稀疏网络,因此学习能力更强,学习样本的规模大. LSM 不仅使用 SVM 的思想,具有良好的分类效果,还具有记忆性,能反映输入序列的历史信息,因此较适合和弦序列识别任务. 目前仅有相关研究将 LSM 用于语音识别[12] ,还未见有研究用于和弦识别领域.

与 LSM 类似的另一种模型是回声状态网络(Echo State Network, ESN) [13] . LSM 和 ESN 提出角度不同且解决的目标问题也不相同, 但被 Verstraeten 等[14] 证明在本质上是一致的. ESN 的中间层回路结构一般使用 Sigmoid 神经元构造,其中间层回路结构使得它也适合解决动态问题,但对音频信号处理能力上不如 LSM.

2.4　和弦序列后处理算法

一般来说,帧长度小于一个和弦的演奏时间,因此,测试音乐中的同一个和弦在输出序列中就会被识别多次. 如被测音乐中一个 Em 和弦的演奏长度是 1s,如果以 600ms 的帧长度,以 400ms 的重复度进行识别,那么这个和弦正确识别序列应该是 6 个 Em: $\{Em,Em,Em,Em,Em,Em\}$. 如果音乐中因为鼓声等噪音而识别错误,如,错误的将序列识别成 $\{Bm,Em,Em,Em,Em,Em\}$ 或 $\{Em,Em,Bm,Em,Em,Em\}$. 对于以上错误,可使用和弦识别序列后处理算法进行修正.

定义奇异和弦为和弦序列中与左右两个相邻和弦都不同的和弦:

1) 第一类奇异和弦. 奇异和弦左右相邻的和弦是相同的;

2) 第二类奇异和弦. 奇异和弦左右相邻的和弦是不同的.

如在和弦序列“CCACFFFBAA……”中,第 3 个和弦 A 属于第一类奇异和弦,第 8 个和弦 B 属于第二类奇异和弦.

基于以下两个规则,设计和弦概率向量以及和弦变换矩阵:

1）不同和弦在音乐中的出现频率是不同的，如最常用的和弦是 C，Dm，Em，F，G，G7，Am；

2）和弦的分布及变换也是有一定规律的，从传统和声的角度来看，大调常以 C 和弦开头，最后以 C 和弦结束；小调则常以 Am 和弦开头，最后以 Am 或 G 结束；C 和弦后面的和弦以 Am 出现的最多.

搜集 100 000 首不同类型音乐的和弦序列进行统计，从中发现 3 031 种和弦. 统计每种和弦出现的次数，为简化模型训练的复杂度，忽略出现率极低的和弦，只保留 500 个和弦. 对向量进行归一化后，便得到一个 500 维的和弦概率向量.

另外，统计和弦转换顺序后，得到一个3 031 × 3 031 的矩阵，矩阵中只保留和弦概率向量中的 500 个和弦，并对矩阵的每行进行归一化处理，得到一个 500 × 500 的和弦变换矩阵，其中第 k 行第 j 列的元素表示第 k 个和弦后面跟的是第 j 个和弦的近似概率.

定义如下 4 个向量：

$\boldsymbol{L}_C$ 代表奇异和弦左边的和弦向量化表示，即奇异和弦左边和弦对应的元素值为 1，其余元素为 0 的 500 维向量；

$\boldsymbol{R}_C$ 代表奇异和弦右面的和弦向量化表示；

$\boldsymbol{C}_{an}$ 代表 LSM 模型输出结果中，概率排名第二的和弦的向量化表示，即该和弦对应的元素值为概率值，其余元素为 0 的 500 维向量；

$\boldsymbol{N}_{ex}$ 代表和弦变换概率矩阵中概率最大的和弦向量化表示，即该和弦对应的元素值为概率值，其余元素为 0 的 500 维向量.

使用如下公式对奇异和弦进行重新估计：

$$\boldsymbol{R}_C = \boldsymbol{L}_C + \boldsymbol{R}_C + \alpha\boldsymbol{C}_{an} + \beta\boldsymbol{N}_{ex},$$

其中，α 和 β 是参数. 最后，奇异和弦的确定值即为向量 $\boldsymbol{R}_C$ 中元素值最大所对应的和弦. 考虑到会有连续出现第二类奇异和弦的情况，因此和弦序列后处理算法是在奇异和弦识别过程中边识别边进行的.

最后，和弦识别结果波形效果图如图 2 所示.

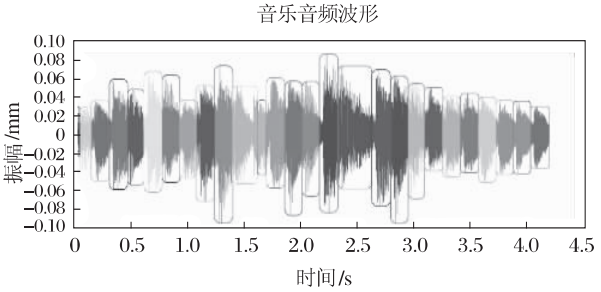

图 2 和弦识别波形效果图

Fig. 2 Result of chord sequence recognition

2.5 总体流程

算法的工作流程如图 3 所示.

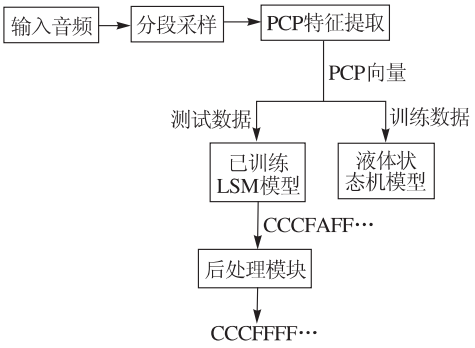

图 3 算法基本工作流程

Fig. 3 Flowchart of the proposed algorithm

3 和弦序列识别实验

3.1 实验设置

和弦概率向量、和弦变换矩阵、和弦转换难度矩阵是先离线统计好的. 实验使用 McGill Billboard 数据集[15]. 该数据集中包含 1958 年～1991 年的 545 首流行音乐，并附带有标注好的音乐和弦序列. 从数据集中随机选出 500 首歌曲，其中 450 首作为训练集，50 首作为测试集. 离线统计数据集包括在网上抓取的 10 000 首音乐和弦谱.

实验使用 Marsyas 工具集（http://marsyas.info/）进行音频切分并提取音乐片段 PCP 特征. 本文使用的液体状态机和回声状态网络开源 Matlab 工具箱由比利时根特大学储备池计算实验室发布. 工具箱中包括液体状态机和回声状态网络的基本实现，包括 Spiking 神经元和模拟神经元，且可设定数据集、设置储备池拓扑结构、交叉验证实验参数. 后处理中的两个参数 α 和 β 分别选为 10 和 20.

实验在相同的训练集上分别训练隐马尔科夫模型、神经网络模型、回声状态网络模型和液体状态机模型，并综合后处理方法，得出实验结果.

3.2 实验与结果分析

实验结果如表 1 所示.

从表 1 中可看出，LSM 方法比 HMM、ANN、ESN 方法具有更高的和弦识别准确性. 对于所有模型，后处理都可提高识别准确率，说明本文提出的后处理算法是有效的.

{C，Dm，Em，F，G，Am，D} 这 7 个和弦是主要和弦，其出现频率在上千个和弦中占到 70% 以上，本

文对方法中主要和弦的识别正确率进行统计，如表 2 所示.

表 1　4 种方法的识别率

Table 1　Recognition rates of 4 methods

%

	HMM	ANN	ESN	LSM
无后处理	63	61	75	77
有后处理	68	70	81	83

表 2　主要和弦的识别率

Table 2　Recognition rates for common chords

%

	C	Dm	Em	F	G	Am	D	综合
HMM	91	92	90	88	91	88	87	90
ANN	90	89	90	88	91	89	87	89
ESN	93	92	92	90	90	89	88	91
LSM	95	92	94	91	92	95	90	93

4　结 束 语

本文首次将液体状态机模型应用于音乐和弦序列识别上，通过与 HMM 方法、ANN 方法、ESN 方法的实验结果对比看出，LSM 具有较好的音乐和弦识别准确率，在 McGill Billboard 数据集上的实验结果达到 83%. 另外本文还提出奇异和弦识别方法及后处理方法，使得识别准确率大幅提升. 本文后处理算法中的两个参数 α 和 β，是根据经验指定的，后续将继续优化该参数.

参 考 文 献

[1] Fujishima T. Realtime Chord Recognition of Musical Sound: A System Using Common Lisp Music // Proc of the International Computer Music Conference. Beijing, China, 1999: 464 – 467

[2] Harte C A, Sandler M B. Automatic Chord Identification Using a Quantized Chromagram // Proc of the 119th Audio Engineering Society Convention. New York, USA, 2005: 6412 – 6419

[3] Ellis D P W, Poliner G E. Identifying Cover Songs with Chroma Features and Dynamic Programming Beat Tracking // Proc of the IEEE International Conference on Acoustics, Speech and Signal Processing. Hawaii, USA, 2007: 1429 – 1432

[4] Ueda Y, Uchiyama Y, Nishimoto T, *et al*. HMM-Based Approach for Automatic Chord Detection Using Refined Acoustic Features // Proc of the IEEE International Conference on Acoustics, Speech and Signal Processing. Dallas, USA, 2010: 5518 – 5521

[5] Harte C, Sandler M, Gasser M. Detecting Harmonic Change in Musical Audio // Proc of the 1st ACM Workshop on Audio and Music Computing Multimedia. Santa Barbara, USA, 2006: 21 – 26

[6] Uchiyama Y, Miyamoto K, Nishimoto T, *et al*. Automatic Chord Detection Using Harmonic Sound Emphasized Chroma from Musical Acoustic Signal // Proc of the International Conference of Music Information Retrieval. Philadelphia, USA, 2008: 901 – 902

[7] Lee K. A System for Acoustic Chord Transcription and Key Extraction from Audio Using Hidden Markov Models Trained on Synthesized Audio. Ph. D Dissertation. Stanford, USA: University, 2008

[8] Lee K, Slaney M. Acoustic Chord Transcription and Key Extraction from Audio Using Key-Dependent HMMs Trained on Synthesized Audio. IEEE Trans on Audio, Speech, and Language Processing, 2008, 16(2): 291 – 301

[9] Sun Jiayin, Li Haifeng, Lei Li. Music Chord Real-Time Perception Based on the Artificial Neural Network // Proc of the National Conference on Man-Machine Speech Communication. Urumqi, China, 2009: 11 – 16

[10] Gerhard D, Zhang Xinglin. Chord Analysis Using Ensemble Constraints // Rás Z W, Wieczorkowska A A, Alicja W, eds. Advances in Music Information Retrieval. Berlin: Springer-Verlag, 2010: 119 – 142

[11] Maass W, Natschläger T, Markram H. Real-Time Computing without Stable States: A New Framework for Neural Computation Based on Perturbations. Neural Computation, 2002, 14(11): 2531 – 2560

[12] Verstraeten D, Schrauwen B, Stroobandt D. Isolated Word Recognition Using a Liquid State Machine // Proc of the 13th European Symposium on Artificial Neural Networks. Brugge, Belgium, 2005: 435 – 440

[13] Jaeger H. The "Echo State" Approach to Analyzing and Training Recurrent Neural Networks with an Erratum Note. Technical Report, GMD148. Bonn, Germany: German National Research Center for Information Technology, 2001

[14] Verstraeten D, Schrauwen B, d´ Haene M, *et al*. An Experimental Unification of Reservoir Computing Methods. Neural Networks, 2007, 20(3): 391 – 403

[15] Burgoyne J A, Wild J, Fujinaga I. An Expert Ground-Truth Set for Audio Chord Recognition and Music Analysis // Proc of the 12th International Society for Music Information Retrieval Conference. Miami, USA, 2011: 633 – 638