

# 长的短时记忆(LSTM)在五声调式和声的应用

贺志强, 冯寅

(厦门大学计算机系 福建 厦门 361005)

【摘要】: 递归神经网络(RNNs)创作歌曲, 但是缺乏全局性、结构不完整, 而长的短时记忆(LSTM)恰恰具有全局性特点。本文以中国五声调式民族乐曲为对象, 修改乐曲输入, 使用改进的 LSTM 为模型, 以不同的方式学习训练, 并最终为已知旋律配和声。

【关键字】: 递归神经网络, 长的短时记忆(LSTM), 和声, 五声调式。

## 1. 引言

和声(harmony)是多声部音乐的音高组织形态, 是音乐的基本表现手段之一。和声通过和弦的运用而表现出不同的音响色彩, 它独立的或同其他因素协同地参与而去表现音乐。和声有结构功能和色彩功能两种基本属性: 结构功能和色彩功能。而结构功能主要表现三个方面: (1)音高纵向结构的组织作用; (2)确立或瓦解调性、调式的作用; (3)发展或终止某一结构的作用。

自动配和声(Automatic Harmonization)系统是一种基于某种和声配置规则及风格, 能生成以已知的旋律为主的多声部音乐的计算机程序。自中世纪以来, 西方就出现了许多多声部音乐。作曲家创作音乐之前, 都需预先精心规划, 以创作出复杂的多声音乐。到了 18 世纪, 就形成了有关音符在纵(多音同时发声)横(下一音符的走势)两方向如何结合的规则体系, 即所谓的和声学。在基础和声学中, 这种同时发生的多音的相互关系, 称为和声。在西方的音乐中, 蓝调由于有规正的和声结构使得它容易在和声理论上有着清晰的解释, 从而有不少系统以蓝调乐曲作为训练集合。例如, Douglas Eck(2002)<sup>[12]</sup>已成功地使用 LSTM 递归神经网络来学习蓝调音乐并能够生成类似风格的新音乐。中国的民间及各民族音乐, 也不都是单声部的。在实际的民间、民族音乐活动中, 无论是民歌、戏曲、曲艺音乐, 还是器乐、歌舞音乐, 都广泛地存在着多声部音乐形式<sup>[19]</sup>。然而中国民族风格化的五声性调式和声理论不及西方传统和声理论那样系统化。

## 2. 现有相关的工作

从 20 世纪, 人们开始使用音乐创作的形式化技术, 使用计算机模拟作曲家的音乐创作过程。现当今计算机作曲使用的技术有: 马尔科夫、神经网络、知识库、遗传算法等。

在早期 80 年代, 相继出现一些使用马尔科夫转换表算法作曲的系统, 算法作曲中按照一个转换表来依次选择音符, 1992 年 Ames 和 Domino[1]使用 Markov 链和随机过程生成了 Cybernetic Compser 系统, 此系统能够创作出诸如像爵士乐、摇摆乐及拉格泰睦乐。此外 Dodge 和 Jerse(1985), Jones(1981), Lorrain(1980)<sup>[2][3][4]</sup>也分别使用马尔科夫转换表算法作了一些作曲的应用。

作为最自然的一种选择是基于音乐知识库的使用。音乐家 David Cope<sup>[5][6]</sup>从事的音乐智能实验(EMI)是注重音乐风格的理解以及复制各作曲家创作风格的课题。Steedman(1984)<sup>[7]</sup>发明了一个生成文法用以描述爵士乐 12 小节蓝调的和弦进程, 而后使用范畴文法, 对其作进一步的精练(Steedman, 1996)<sup>[8]</sup>。

另外人工神经网络也被广泛地使用在音乐应用系统中。Mozer<sup>[9]</sup>使用递归神经网络技术构造了 CONCERT, 并用反向传播学习算法来训练 CONCERT, CONCERT 能够以一音符一音符的方式来创作旋律。但是其生成的旋律缺乏音乐的全局关联性, 即它无法获取较高级的音乐特征, 这和递归神经网络中学习梯度的逐渐消失有关(Hochreiter 和 Schmidhuber, 1997)<sup>[10]</sup>。一种被称之为长的短时记忆 LSTM 递归神经网络可以解决这一问题(Felix A. Gers 和 Jurgen Schmidhuber, 2001)<sup>[11]</sup>。Douglas Eck(2002)<sup>[12]</sup>已成功地使用 LSTM 递归神经网络来学习蓝调音乐并能够生成类似风格的新音乐。

本文试图以五声调式的中国民族音乐对象, 讲述了如何使用长的短时记忆(LSTM)为旋律配和声。

## 3. 问题的提出

在使用 Douglas Eck 中的 LSTM 模型应用到中国五声调式乐曲中, 由于和弦上的差异, 我们不能得到很好的结果。而且 Douglas Eck 中过于规正严格(只能使用八份音符作为最小的处理单位), 而实际上的民间、民族音乐活动中, 无论是民歌、戏曲、曲艺音乐, 还是器乐、歌舞音乐, 都广泛地存在着多声部音乐形式<sup>[15]</sup>。且中国民间及各民族多声部音乐, 没有什么完整的和声体系。目前, 虽然出现了不少阐述中国民族风格化的五声性调式和声理论的书籍和论文<sup>[16][19][17]</sup>, 但它们似乎不及西方传统和声理论那样系统化。所以我们需要对 Douglas Eck 有关蓝调创作使用的 LSTM 学习及训练过程加以改进。

## 4. 长的短时记忆(LSTM)模型

### 4.1 长的短时记忆(LSTM)

由于篇幅限制, 本文就不深入详细介绍长的短时记忆(LSTM), 详细情况可参考<sup>[13]</sup>中所讲述的。简言之, 设计 LSTM 是为了得到不随时间变化而改变的误差以及防止这个误差不会被干扰。LSTM 利用线性单元克服误差消失和误差的指数增长问题, 每个单元有一个固有的自我环绕的回路, 并且单元周围有许多非线性的门单元控制信息流。同样, 在 LSTM 中训练过程采用改进的 BPIT 和 RTRL 梯度下降算法。

### 4.2 程序中音符形式描述

把简谱或则五线谱描述的乐曲输入程序实验中, 需要把乐曲中的音符转化为计算机可以处理的数据格式。为此, 我们把实验中的音符的属性描述为三维向量  $\langle N, S, T \rangle$ , 其中  $N$  是音符名集合  $\{rC_m, rD_m, rE_m, rF_m, rG_m, rA_m, rB_m\}$  中一个元素。这里  $m(=0, 1, 2, \dots)$  代表音符所属音组 ( $m$  相同表示同属一个八度内),  $r$  为半音升降号 ( $\#, b$ ) 或为空。  $T(=1$  (全音符),  $=2$  (2 分音符),  $=4$  (4 分音符),  $=8$  (8 分音符),  $=.1$  (附点全音符),  $=.2$  (附点 2 分音符),  $=.4$  (附点 4 分音符),  $=.8$  (附点 8 分音符), ...) 等为音符的时值描述。

### 4.3 五声调式和弦字典库

如同基于规则作曲系统, 需预先定义一些文法规则, 而在本文中我们需预建立不同调式所有的和弦字典库。和弦字典库由一系列和弦词条组成。我们定义一个和弦词条描述如下:

和弦音表 (C, C <sub>2</sub> ... C <sub>n</sub> )	和弦标识
--	------

该字典库是人工分析得到的, 根据文献<sup>[16]</sup>所指出的调式和弦的可能结构, 表(一)仅给出宫调式的所有可能的主功能、属功能、及下属功能的和弦词条。在此表格中 T11-T15 属于主和弦类, DV11-DV12 为属和弦类中第 1 类, DIII1-DIII3 为属和弦类中第 2 类, DVII1-DVII5 是属和弦类中第 3 类, SV11-SV6 是下属和弦类中第 1 类, SII1-SII3 是下属和弦类中第 2 类, SVII1-SVII5 是下属和弦类中第 3 类

## 5. 实验及结果

### 5.1 学习过程

本实验中我们可以测试 LSTM 学习复制宫调式民族歌曲和声整体结构的能力。输入到 LSTM 网络中的数据形式以和声为乐

曲结构,这样可以确保模型对长时间延迟结构的灵活性,每个处理单元以节奏为准.在网络中设置 4 个组块,并且每个组块包含 2 个细胞单元.4 个组块相互连接,并且每个组块与输入层连通,输出层与每个细胞单元、输入层连通.遗忘门、输入门、输出门在 4 个组块上的的阈值分别为: -0.1、-0.3、-0.5、-0.7,而输出阈值为 0.5,学习率为 0.00001,激励率为 0.9,权值每步都要更新.

5.1.1 乐曲与和弦标志间的转换

当把乐曲输入到程序之前,我们需要首先对乐曲音符中的装饰音进行处理及选择学习训练的最小单元,即以什么情况为一个学习单元(2/4 与 4/4, 3/4 与 3/8、6/8 分别做相同处理).对装饰音本文不予考虑,即把该音从乐曲中删除.而对于不同节奏的情况,当把乐曲输入,我们以 2/4 为例,程序一个节拍为一个处理单元,并且在一个节拍中的音符中找出一个和弦,使得一个节拍的音符在该和弦中尽可能多出现.经过处理后,带有和弦标志的和弦序列作为新的乐曲输入,经过学习训练得到乐曲的和声整体变化特性.图 1 为宫调式《依马亚吉松》藏族民歌[18]乐曲的一部分.

序号	和弦标志	和弦标识	序号	和弦标志	和弦标识
1	(C0 C1)	T1 <sub>1</sub>	20	(B0 D1 #F1)	DV1 <sub>1</sub>
2	(C0 G0 C1)	T1 <sub>2</sub>	21	(#B0 D1 F1 A1)	DV1 <sub>2</sub>
3	(C0 B0 G0 C1)	T1 <sub>3</sub>	22	(B0 D1 #F1 A1)	DV1 <sub>2</sub>
4	(C0 B0 G0 A0 C1)	T1 <sub>4</sub>	23	(#B0 D1 F1 G1)	DV1 <sub>1</sub>
5	(C0 G0 A0 C1)	T1 <sub>5</sub>	24	(B0 D1 F1 A1)	DV1 <sub>1</sub>
6	(G0 D1)	DV1 <sub>1</sub>	25	(F0 A0 C1 D1)	S1 <sub>1</sub>
7	(G0 C1 B1)	DV1 <sub>2</sub>	26	(F0 C1 D1)	S1 <sub>2</sub>
8	(G0 B0 D1 E1)	DV1 <sub>3</sub>	27	(F0 A0 C1)	S1 <sub>2</sub>
9	(G0 D1 A1)	DV1 <sub>4</sub>	28	(F0 A0 C1 E1)	S1 <sub>2</sub>
10	(G0 D1 F1 A1)	DV1 <sub>5</sub>	29	(F0 G0 C1 D1)	S1 <sub>2</sub>
11	(G0 C1 D1 A1)	DV1 <sub>6</sub>	30	(F0 G0 C1)	S1 <sub>2</sub>
12	(G0 D1 E1 A1)	DV1 <sub>7</sub>	31	(D0 F0 A0)	S1 <sub>2</sub>
13	(G0 B0 D1)	DV1 <sub>8</sub>	32	(D0 A0 C1)	S1 <sub>2</sub>
14	(G0 B0 D1 A1)	DV1 <sub>9</sub>	33	(D0 F0 A0 C1)	S1 <sub>2</sub>
15	(G0 B0 D1 F1)	DV1 <sub>10</sub>	34	(A0 C1 E1)	S1 <sub>2</sub>
16	(G0 B0 D1 F1)	DV1 <sub>11</sub>	35	(A0 E1)	S1 <sub>2</sub>
17	(E0 G0)	D1 <sub>1</sub>	36	(A0 C1 E1 G1)	S1 <sub>2</sub>
18	(B0 G0 D1)	D1 <sub>2</sub>	37	(A0 C1 G1)	S1 <sub>2</sub>
19	(E0 G0 B0 D1)	D1 <sub>3</sub>	38	(A0 E1 G1)	S1 <sub>2</sub>

表(一):宫调式的各种可能的主功能、属功能、及下属功能的和弦词条



图 1:宫调式《依马亚吉松》前 18 个小节

经过处理转换得到的新的序列为: T14 D1111 | T12 S1V5 | D1112 D1112 | D1112 D1112 | DV1 DV115|SV11 SV12 | T12 T14 | DV4 DV5 | DV4 D1113|DV3 DV9 | DV4 S13 |T12 T14| SV11 T14 |T15 T15 | S1V6 SV1| DV5 SV2 | T15 T15 |DV5 DV5.把新得到的序列及旋律乐段一同输入到 LSTM,经过学习训练得到该乐曲的和声整体变化规律及旋律音符串与和声之间的概率.



图 2 江西民歌《八月桂花遍地开》旋律乐曲配的伴奏乐曲

5.2 配和声

该小节我们以宫调式的 2/4 节奏的江西民歌《八月桂花遍地开》为例.为旋律乐曲配和声,输入新的旋律乐曲,由 LSTM 得到该乐曲的和声整体变化即得到该乐曲的和弦标识串.此外对于旋律乐曲的每个节拍中的音符串,在其和弦标识所在和弦类中匹配其最大相似旋律音符串,把对应音符串的伴奏音作为该旋律乐曲的伴奏音.图 2 为为宫调式的 2/4 节奏的江西民歌《八月桂花遍地开》配的伴奏乐曲.

5.3 算法流程

学习过程:

乐曲输入

音符转换及建立和弦字典库

LSTM 学习训练得到和弦变化特性及旋律音符与和弦之间的概率

配和声过程:

旋律乐曲输入

音符转换并生成和弦标识序列

对每个节拍中的音符串查找其极大相似的乐曲旋律音符串拷贝伴奏乐曲音符串

6. 结论

本文基于 LSTM 的神经网络很好地学习了音乐的整体结构,并能用这些学到的知识对已知旋律配出新的和声,和声很好地起到渲染、衬托的作用,使音乐形象更加深刻、鲜明.此外由于中国民族与蓝调音乐上的和弦上的差异,本文很好的解决了中国民族调式乐曲对 Douglas Eck [14] 的不适应性,而且解决了 Douglas Eck 仅仅以八分音符为最小度量的缺点,本文可以选择不同的学习风格选择处理音符串的组成.然而本文的 LSTM 使用了音乐中和声还是有很大地方需要改进提.首先人工构建的和弦字典库可能不能包含所有合理的和弦.另外我们的系统中没有处理装饰音、在乐曲的转调、和弦的转位等等上处理不足,使得生成的和声不太丰富,过去单调.最后我们为旋律乐曲生成的和弦依赖于选择的训练乐曲,因为我们生成的和弦是从训练乐曲中选择算法认为合理的和弦,造成伴奏上的过去机械化、缺乏创造性.

参考文献:

1. C.Ames and M.Domino. Cybernetic composer: An overview. In M.Balaban, K.Ebcioğlu, and O.Laske, editors, Understanding Music with AI. AAAI Press, 1992. 186-205.
2. Dodge C.and Jerse T.A. Computer Music: Synthesis, composition, and performance. New York:Shirmer Books, 1985.
3. Jnes K. Compositional applications of stochastic processes. Computer Music Journal, 1981, 5(1):45-61.
4. Lorrain D. A panoply of stochastic 'cannons'. Computer Music Journal, 1980, 4(3):48-55.
5. David Cope. Computer Modeling of Musical Intelligence in EMI. Computer Music Journal, 1992, 16(2):69-83.
6. David Cope. Experiments in Musical Intelligence. Madison, WI:A- R Editions, 1996.
7. M. Steedman. A Generative Grammar for Jazz Chord Sequences Music Perception, 1984, 2:52-77.
8. M. Steedman. The Blues and Abstract Truth: Music and Mental Models. In A.Garnham and JOakhill, editors, Mental Models in Cognitive Science. Erlbaum, Mahwah, NJ, 1996.
9. M.C. Mozer. Neural network composition by prediction: Exploring the benefits of psychophysical constraints and multiscale processing. Cognitive Science, 1994, 6:247-280.
10. S. Hochreiter and J Schmidhuber. Long Short- Term Memory. Neural Computation. 1997, 9(8):1735-1780.
11. F. A. Gers and J Schmidhuber. LSTM recurrent networks learn simple context free and context sensitive languages. IEEE Transactions on Neural Networks, 2001, 12(6):1333-1340.
12. Douglas Eck. Finding Temporal Structure in Music: Blues Improvisation with LSTM Recurrent Networks. In H.Boulard editor, Neural Networks for Signal Processing XII, Proc. of the 2002 IEEE Workshop, New York, IEEE 2002. 747-757.
13. A. Gers and JSchmidhuber. LSTM recurrent networks learn simple context free and context sensitive languages. IEEE Transactions on Neural Networks, 12(6):1333- 1340,2001.
14. Douglas Eck. Finding Temporal Structure in Music:Blues Improvisation with LSTM Recurrent Networks. In H.Boulard editor,(下转第 136 页)

过此模块可以快速看到各案件的项目名称、案件编号、委托单位、要求完成时间以及目前进展状态, 通过点击还可显示各项中各个操作进程的详细信息。

(2) 房产收件: 自动显示已经收件但还未进行任务分配的项目信息, 通过此模块可以对接收的案件信息进行登记和修改。项目收件信息包含客户单位信息、项目相关信息以及案件收件的详细信息。客户单位信息包含单位名称、地址、联系人、联系电话; 项目相关信息包含项目名称、项目规模、测绘类型、坐落地址、收费价格、预估面积等; 案件收件信息包含资料名称、资料类型、数量和接收时间等。

(3) 任务分配: 自动显示项目任务分配信息, 通过此模块可以对项目进行任务分配同时可以查看任务的相关分配信息。任务分配信息包含任务安排时间、安排人、项目负责人、测绘日期以及每个作业人员的姓名、作业内容。

(4) 外业测绘: 自动显示项目外业测绘信息。当作业人员外业测绘结束后登记其外业测绘情况信息。项目负责人和相关管理者通过此模块可以快速了解到每个项目和每个作业员的作业进展情况。外业测绘信息包含作业人员、作业时间和作业内容。

(5) 内业整理: 自动显示项目内业整理信息。当作业人员完成外业测绘并对数据进行内业整理后, 可对其分配到的任务进行内业整理信息登记。项目负责人和相关管理者可以根据每个作业员的测量面积统计出项目总面积。内业整理信息包含作业人员、作业时间、作业内容、测量面积以及碰到的相关问题反馈信息。

(6) 技术总结: 一般是由项目负责人对案件的总体情况包括测绘中发现的问题进行总结。技术总结信息包含总结人员、总结时间、总结内容、客户对本项目的技术总结的反馈意见。

(7) 质量检查: 质量检查包含一检和二检, 质检员通过此模块登记质检情况信息, 作业人员和相关领导人员可以了解到每个项目和每个作业人员的作业质量情况。质量检查信息包含检查人员、检查时间、检查类型、检查内容、检查意见和检查结果。

(8) 出具成果: 登记出具给房地产行政管理部门验收审核的已完成项目的资料信息。出具成果信息包含出具时间、出具资料类型、名称和数量单位等。

(9) 案件跟踪: 对出具给房地产行政管理部门验收审核的案件通过, 退改情况进行登记。

(10) 成果归档: 将项目相关材料和测绘成果数据进行归档, 归档信息通过文件上传来实现, 同时相关人员还可以对归档信息进行下载和浏览。成果归档信息包含资料名称、归档时间、归档人员和数量单位。

(11) 质量追踪: 对项目的质量满意度情况进行调查。质量追踪信息包含调查时间、投诉处理时间、顾客意见和顾客满意度。

(12) 财务监控: 对承接每宗项目的测量费用进行自动登记, 包括应收款项、已收金额、未收金额等。

3.3.4 拆迁测量: 为了满足新农村规划和旧城改造拆迁安置的业务发展需要, 特增添此模块。本模块含有项目概括、作业依据、生产进度、质量检查、出具成果登记、验收归档和项目工作总结等内容。

3.3.5 生产统计: 相对于系统所涉及的生产项目进行电脑自

动化统计。包含生产任务统计、工作量化统计、生产产值统计。生产任务统计是对项目总数量、房产测绘项目、拆迁测量项目、已完成项目、正在生产项目、超时完成项目等内容进行统计。工作量化统计, 主要是作业员的个人工作量以及具体某项目的测绘面积总量统计。生产产值统计, 是指整个单位的所有测量项目收入进行统计, 为财务监控和年度产值汇总提供数据来源。

3.3.6 人力资源管理: 为了适应现代人力资源管理的发展趋势, 设计此模块。目前已有员工档案、培训申请、培训计划、培训档案、培训记录内容, 也可以根据需要适时增添一些人力资源管理其他方面的专题, 如人力规划、激励、考核等。

3.3.7 后勤办公管理: 包含设备采购、设备登记、设备借用、设备报废, 印章使用及其他的现代行政办公管理方面的内容。

3.3.8 政策法规: 主要是为了满足终端客户查看与房地产测绘有关的法规和条例, 如《中华人民共和国国家标准房产测量规范》、《福州市房屋面积技术细则(暂行)》等, 以及单位内部的作业规定, 可按标题、上传时间等条件进行查询。

3.3.9 资料下载专区: 为系统功能模块外的资料提供共享平台。各用户可以上传、下载与房产测绘相关的技术见解、业务探讨、作业技巧等内容。

#### 4. 系统实现

系统采用开发语言使用 Java, JSP, JavaScript, Java 集成开发环境采用 Bland 公司的 JBuilder 9, JavaScript 编辑工具采用 Editplus2.1 和 UltraEdit-32, WEB 页面制作采用 DreamWeaver MX 和 Macromedia Fireworks MX 2004, 团队文档管理和版本维护采用 Visual Source Safe 2002, 建模语言软件采用 Rational Rose Enterprise Edition 2003。系统开发界面如图 2, 主要由三部分组成: 系统上方为功能菜单栏; 左边为功能区, 点击某个具体的功能按键, 即可显示其对应的内容; 中间为功能的具体操作区与列表, 根据不同的功能模块特点, 可以增加、删除、修改、查询内容的选择。

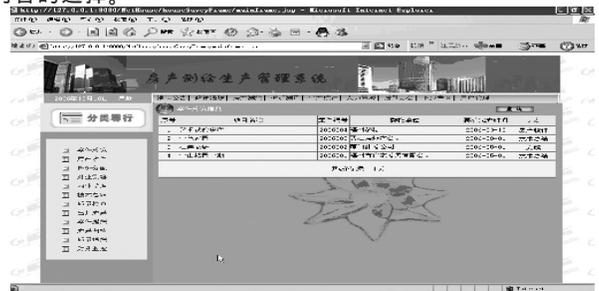


图 2 系统界面图

#### 5. 结束语

随着房产测绘事业的发展, 网络技术和数据库技术的深入应用, 开发一套满足房产测绘管理及应用的信息系统是非常必要的。本文从系统分析、功能设计和系统实现等方面提出了房产测绘生产管理系统的的设计思路和实现方法。该系统的实现和应用, 有利于推动房产测绘生产管理的规范化, 可进一步提高生产效率和信息化水平, 并为领导的决策提供科学依据。

#### 参考文献:

1. 张建强《房地产测量》测绘出版社

18. 隆翔编著, 中央民族大学出版社, 中国少数民族钢琴通俗曲及调式基础练习指法, 2003.12 P30- P31.  
19. 冯寅 周昌乐 算法作曲的研究进展,《软件学报》, Vol.17, No.2(2006), 209- 215.

(上接第 102 页)

Neural Networks for Signal Processing XII, Proc. of the 2002 IEEE Workshop, New York, IEEE 2002. 747-757.

15. 樊祖荫:《中国五声性调查核实式和声的理论与方法》, 上海音乐出版社, 2003 年 12 月。

16. 张肖虎:《五声性调式及和声手法》, 人民音乐出版社, 1987 年 12 月。

17. 郭瑛等:《实用和声学简明教程》, 南京师范大学出版社, 1999 年 10