### MC, HMM and LSTM for generation of music

#### Markov Chain in Music Generation

假设一首乐曲可以由一系列的随机变量(i.e.一个随机向量)表示, 马尔科夫链(Markov Chain)[ref…]生成音乐的基本假设是, 这些随机变量遵循马尔可夫性质, 即

并且, 假设这些随机变量的边缘分布同分布

那么我们可以根据观察到的数据得到平稳分布的估计

其中是对观测数据中音符i出现的频率, 以及对转移矩阵的估计

其中是音乐序列中出现的次数. 由此, 我们可以通过平稳分布和概率转移矩阵得到序列联合分布

若想要从其中采样出一个序列, 我们可以简单地的通过贪心法来取到一个合理的序列

也可以采用简单的采样法来取得一个序列

我们也可以最大化这些序列的对数似然

可以通过动态规划算法求出确定性的最优解, 其状态转移方程如下

这里代表取一个长为i的序列, 且结尾音符为j时最大的对数似然, 时间复杂度是. 除了以上列举的一些简单方法, 其实还有一种有趣的采样方法, 就是Gibbs采样, 通过将Markov链绕成一个环, 我们可以得到对称的条件概率模式

于是我们可以通过Gibbs采样来得到一个概率较高的样本. 概括的来讲,Gibbs采样每次固定除了一个分量之外的其它分量, 并根据条件概率分布对没有固定的分量进行采样, 然后轮流对每一个分量进行如此操作, 可以证明, 如此进行足够多的轮数之后, 序列收敛于联合分布(即整个Markov Chain)的一个最概然序列. 值得一提的是Gibbs采样也广泛用于大量其它种类的概率分布的采样, 衍生出了MCMC等概率模型中的重要方法.

|  |
| --- |
| Gibbs Sampling Algorithm |
| 1. |
| 2. For i = 1…k, |

#### HMM: Hidden States in Music

可以想见, 乐曲并不完全遵循Markov性质, 一首乐曲的某个音符很难只由其上一个音符决定, 而往往由上一段乐曲, 以及整段乐曲及其部分的语义决定. 隐式马尔可夫模型(HMM, Hidden Markov Model)指出, 一段观察序列可能由隐含状态

所决定, 其中隐含状态又是符合Markov性质转移的, 某种意义上代表着序列某时刻的语义信息(比如此时应该时舒缓还是激烈, 欢快或是忧伤), 同样具有转移矩阵

以及输出矩阵, 决定了从隐含状态到观察状态的转移概率

其中是音乐序列中出现的次数. 不过一般来说, 我们难以得知隐含状态的序列, 而往往只知道观察状态的序列, 这时候我们就需要利用Baum-Welch算法来估计一个转移矩阵(以及输出矩阵), 其对应了最大的观察序列的后验对数似然. Baum-Welch算法是EM算法的一个应用, 这里我们将不介绍他的推导, 而是直接给出算法

|  |
| --- |
| Baum-Welch Algorithm[ref wikipedia…] |
| 通过前向算法计算前向概率:1. 2. |
| 通过后向算法计算后向概率:1. 2. |
| 计算中间变量 |
| 更新平稳分布,转移矩阵和输出矩阵 |

得到最优的转移矩阵和输出矩阵后, 也可以和显式马尔科夫模型一样通过贪心, 直接采样, 动态规划或者Gibbs采样的方法求出一些较好的序列.

#### LSTM: Recurrent Networks in Music Generation

HMM部分解决了乐曲隐含语义的问题, 将其编码在了模型的隐含状态中, 但是没有解决乐曲的另一个特性: 一个音符往往不只是和前一个时刻的状态有关, 而是和之前数个甚至是整段乐曲有关. 递归神经网络利用时序传播特性完美编码了这种特性, 其中最实用和有趣的模型就是长短时记忆网络(LSTM, Long Short Term Memory Network)[ref…]. 简单地说, 他会接受一个序列数据作为输入, 并且按照时间顺序依次将每个时间步的输入和之前计算得到的长期和短期记忆的隐含状态通过自己的内部计算单元(LSTM单元,包含输入门, 输出门和最重要的记忆/遗忘门, 如下图)得到新的隐含状态, 然后通过新得到的隐含状态, 类似HMM那样得出一个输出状态. 这样, LSTM就可以完成对乐曲局部和全局语义信息的“记忆”, 克服MC模型的无记忆性, 以及HMM模型没有解决的全局-局部语义问题! 这里我们使用序列部分预测的方式, 来对LSTM模型进行训练.



#### Experiment

我们主要采用的是POP909流行音乐数据集, 包含了909段比较著名的流行乐旋律, 用MIDI文件格式保存. 我们从这些MIDI文件中提取出了包括休止符在内的旋律序列, 并且以此为基础生成音乐. 在最基本的马尔科夫链生成音乐中, 我们使用一首乐曲或者数首乐曲融合统计的方式得到马尔科夫链的转移矩阵, 并且采用上述的几种采样方法(贪心法, 直接采样法, 动态规划和Gibbs采样)来得到生成的采样序列. 对于HMM和LSTM方法, 我们从数据集POP909数据集中随机抽取并且随机剪裁了50000个音乐片段, 之后使用它们对模型进行参数优化, 最后通过动态规划和Gibbs采样法得到生成的音乐, 对于LSTM来说, 具体的优化任务是通过已知的序列来预测未知的下一个音符. 完整的代码可以参见[ref: github], 其中数据读取和MC模型使用Python编写, HMM和LSTM使用Mathematica编写. [此处应该加入三种的比较…就简单放三段乐谱凑页数]