

人工智慧與深度學習

期中論文報告

D0745378 薛竣祐

D0827036 吉 偉

D0987486 曾翊昇

Bach Style Music Authoring System based on Deep Learning

基於深度學習的巴赫風格音樂創作系統

介紹論文



因為技術發展 傳統音樂發展迅速 Johann Sebastian Bach (1685-1750)

巴赫的音樂具有明顯的樂理特徵，音程和節奏的規律性強，有利於神經網絡提取特徵。在算法組成和人工智能組成中，研究人員的研究核心是如何讓計算機理解並找出概率分佈以及復雜音樂背後的數字邏輯。

LSTM、巴赫

LSTM(Long Short Term Memory)

使用LSTM學習巴赫音樂的深度神經網絡模型。這是第一個將神經網絡應用於巴赫風格音樂創作的研究

訓練規則如下:通過給定長度的音符和時長序列預測下一個音符和時長的值,與訓練集中的期望值進行比較,修改 LSTM 權重以通過損失函數減少錯誤並獲得更好的結果

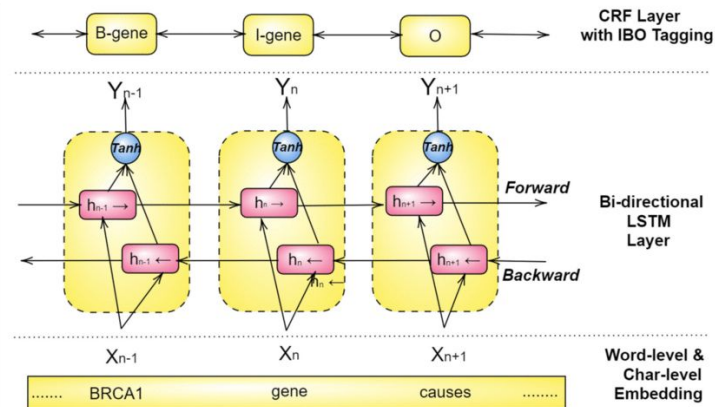
音符序列的訓練使用Bi-LSTM層來考慮前向和後向信息,使預測更加準確。

演算法的選擇

HMM(1958): 隱馬爾可夫模型，雖然模型效果良好，但生成的結果缺乏多樣性，僅能將樣本片段重新組合。

GAN(2018): 生成對抗網路的模型訓練難以收斂，儘管收斂也容易崩潰。其模型訓練成本太高且需耗時許久。

Bi - LSTM: 為了確保音高、音符時值，都能準確的預測，所以決定透過正向LSTM與反向LSTM的結合，比起傳統STM能更精準的生成文本，能控制網路複雜度並漸少損失率。



數據處理流程

1. 引入MUSIC21資料
2. 將音樂和弦分離成個別音符
3. 將音符拆分為音高、音符時值資料
4. 標準化資料
5. 切割成資料序列

$$X = \begin{bmatrix} note_1 & note_2 & note_3 & & note_n \\ note_2 & note_3 & note_4 & \dots & note_{n+1} \\ note_3 & note_4 & note_5 & & note_{n+2} \\ & \vdots & & \ddots & \vdots \\ note_m & note_{m+1} & note_{m+2} & \dots & note_{m+n} \end{bmatrix} \quad R = \begin{bmatrix} note_{n+1} \\ note_{n+2} \\ note_{n+3} \\ \vdots \\ note_{n+m+1} \end{bmatrix}$$

Figure 1 Input and Output Note Sets

數據處理困難

1. 歌曲斷點、破壞模型訓練穩定度
2. 訓練模型架構層設計

訓練

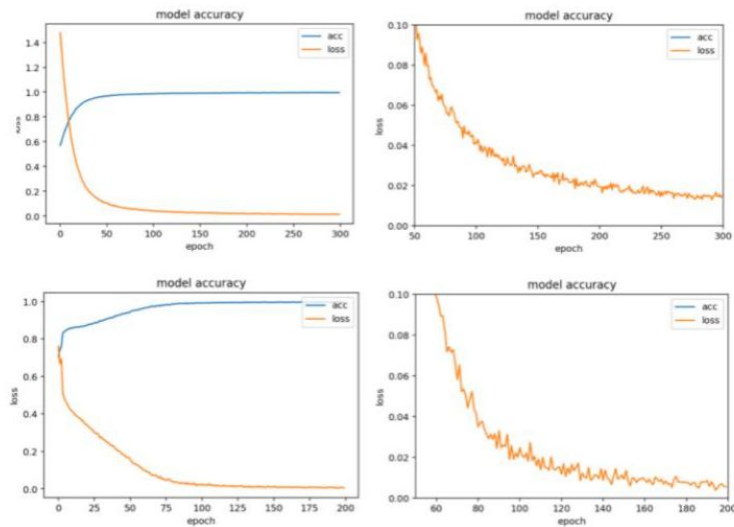
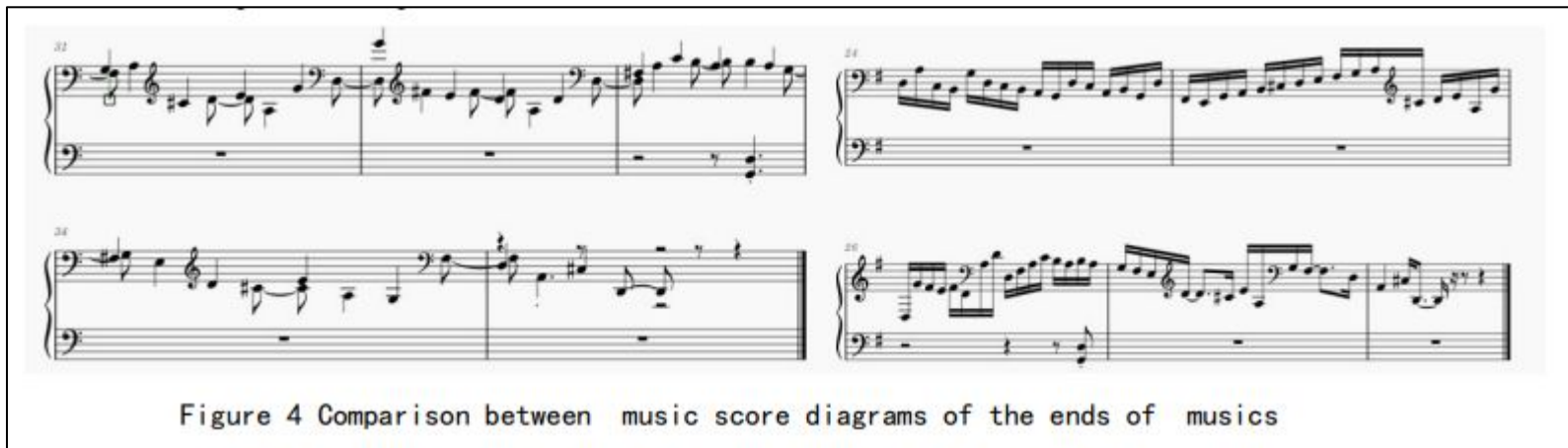


Figure 2 Model Training of loss and accuracy values

結果



結果

Table 3 Music Evaluation Scores

number	Order	Author	Interval Score	rhythm Score	Melody Score	Overall Score
M1	1	Bach	4.0	4.7	3.8	4.17
M2	2	LSTM	3.8	4.3	3.7	3.93
M3	6	Bach	3.3	3.5	3.4	3.40
M4	4	LSTM	3.8	4.0	3.8	3.87
M5	5	LSTM	4.0	3.8	3.5	3.77
M6	2	Bach	3.8	4.2	3.8	3.93

結果

Table4 Distinguishing Result

	M1	M2	M3	M4	M5	M6	Not sure
Distinguish ed as the music composed by AI	15.79%	21.05%	26.32%	21.05%	26.32%	36.84%	15.79%

Table 5 familiar Music

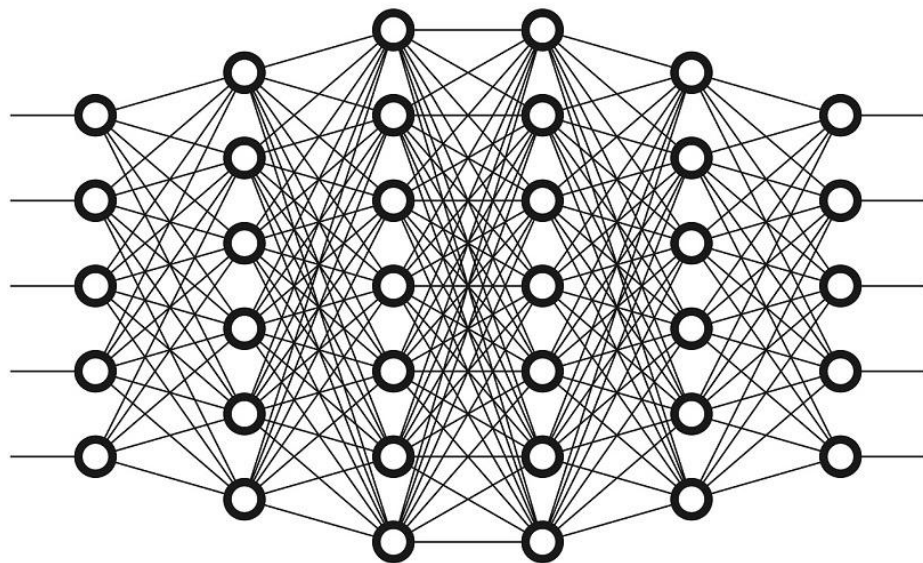
	M1	M2	M3	M4	M5	M6	Have not heard
Having heard the music before test	21.05%	0.00%	0.00%	0.00%	0.00%	10.53%	73.68%



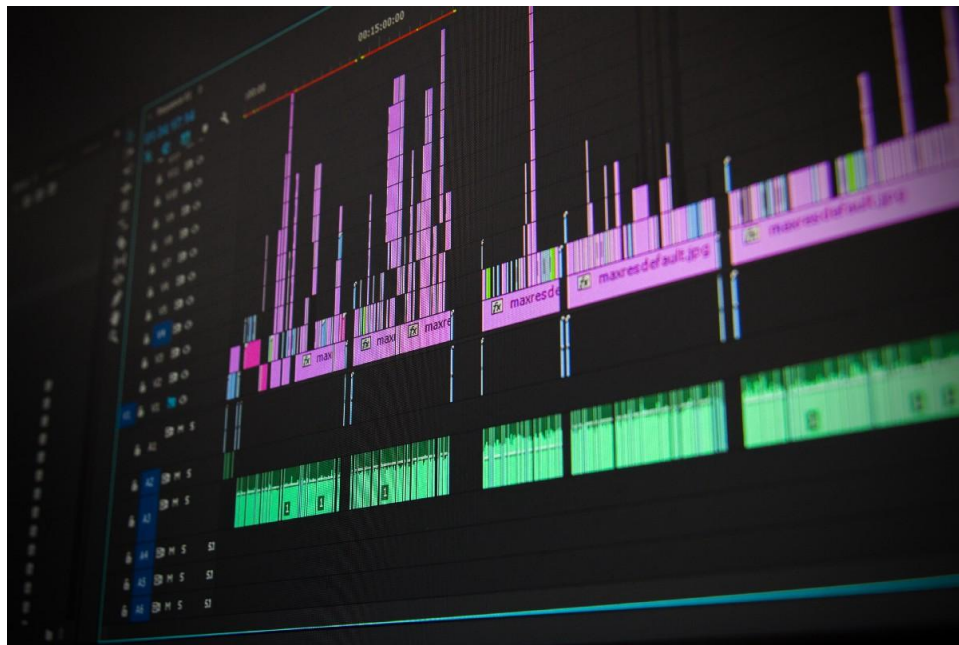
Novice-AI Music Co-Creation via AI-Steering Tools for Deep Generative Models



摘要

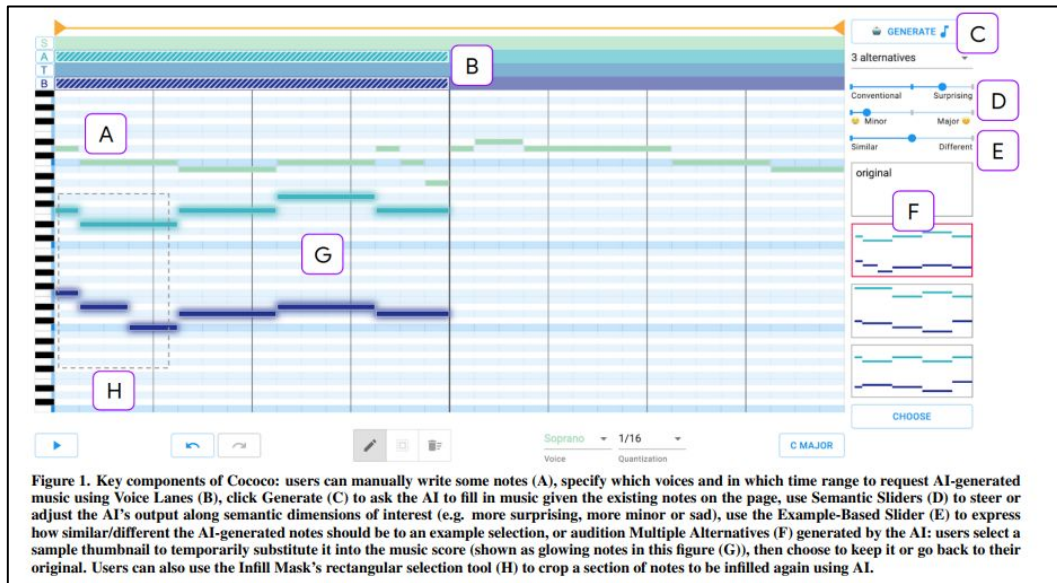


介紹

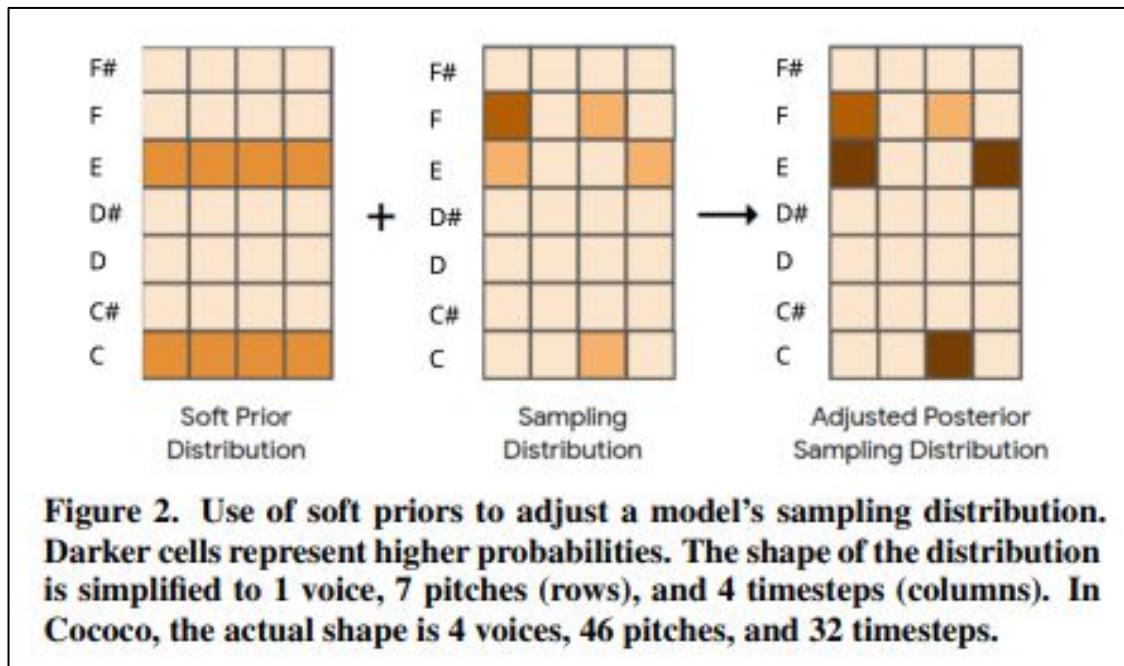


介面與功能

1. Voice Lanes
2. Semantic Sliders
3. Audition Multiple Alternatives
4. Example-based Slider
5. Soft Priors



介面與功能



結果

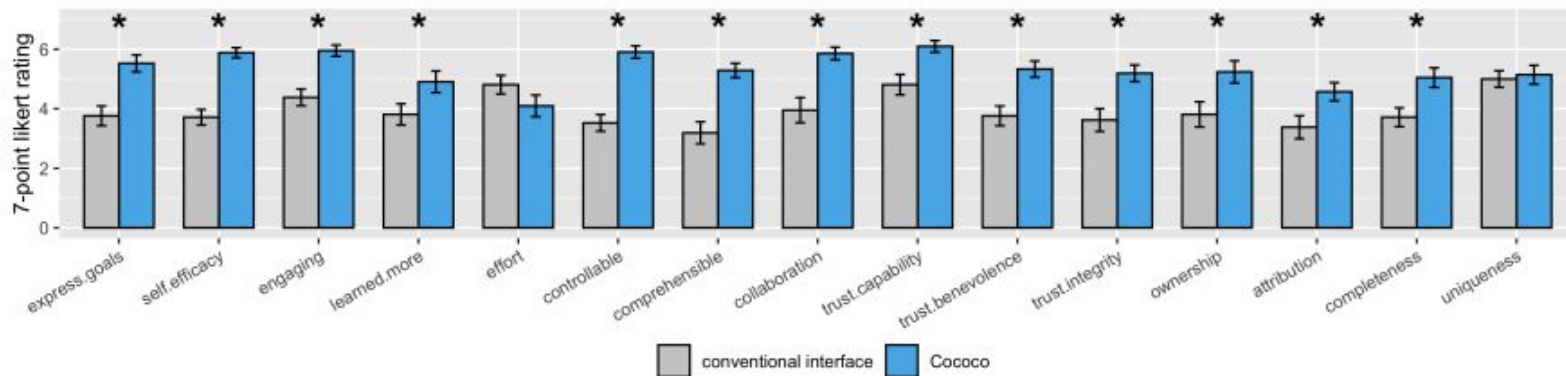




Figure 3. Results from post-study survey comparing the conventional interface and Cococo, with standard error bars.



Predicting the Preference for Sad Music: The Role of Gender, Personality, and Audio Features

預測對悲傷音樂的偏好：性別、個性和音頻特徵的作用



論文介紹

悲劇的感官影響，能喚醒人們情感，引發吸收、沉思、幻想等思緒。因此對於悲劇可能會有反常理的喜好，這便是悲劇悖論。

而本篇旨在研究及預測悲劇悖論對於性別及大五人格因素產生音樂偏好的影響。

未來可發展為音樂推薦系統演算法

研究方法

隨機挑選60首流行音樂，並擷取其25秒片段，請約100人為歌曲標記1~5分的悲傷程度，也為每個受測者進行性向測驗。

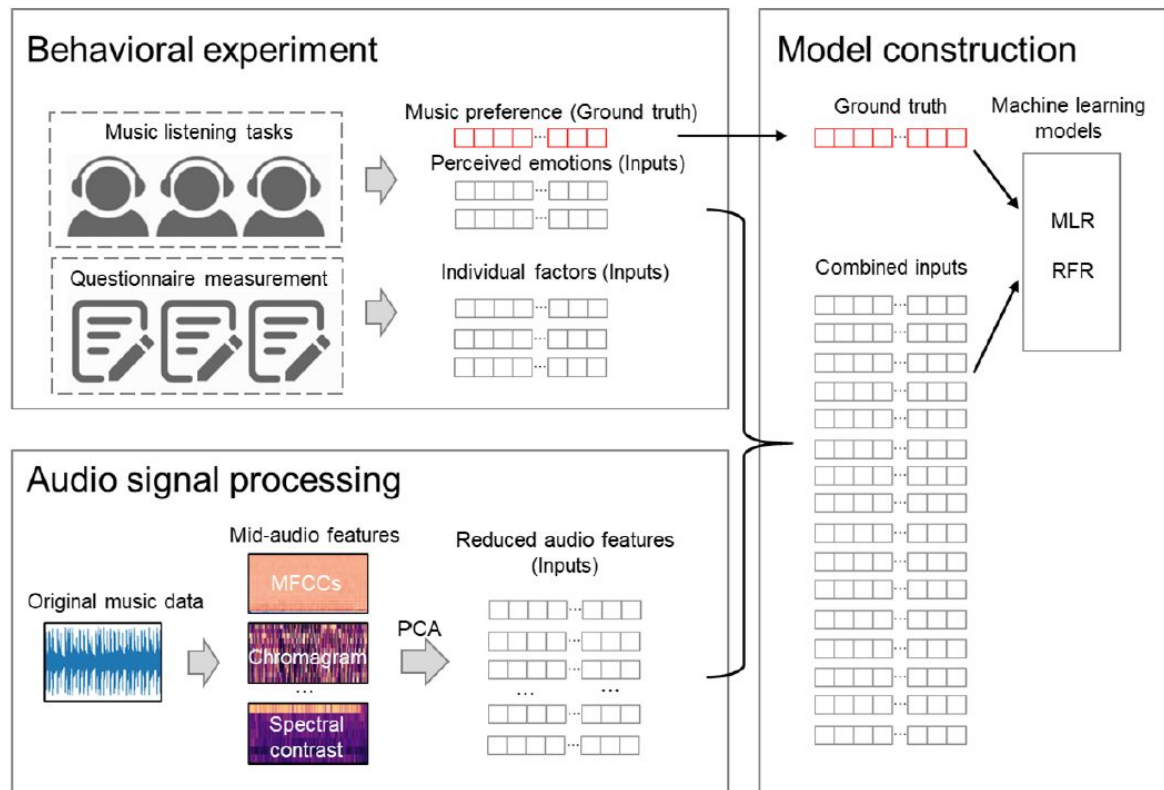


FIGURE 1. The framework of this study.

模型建構

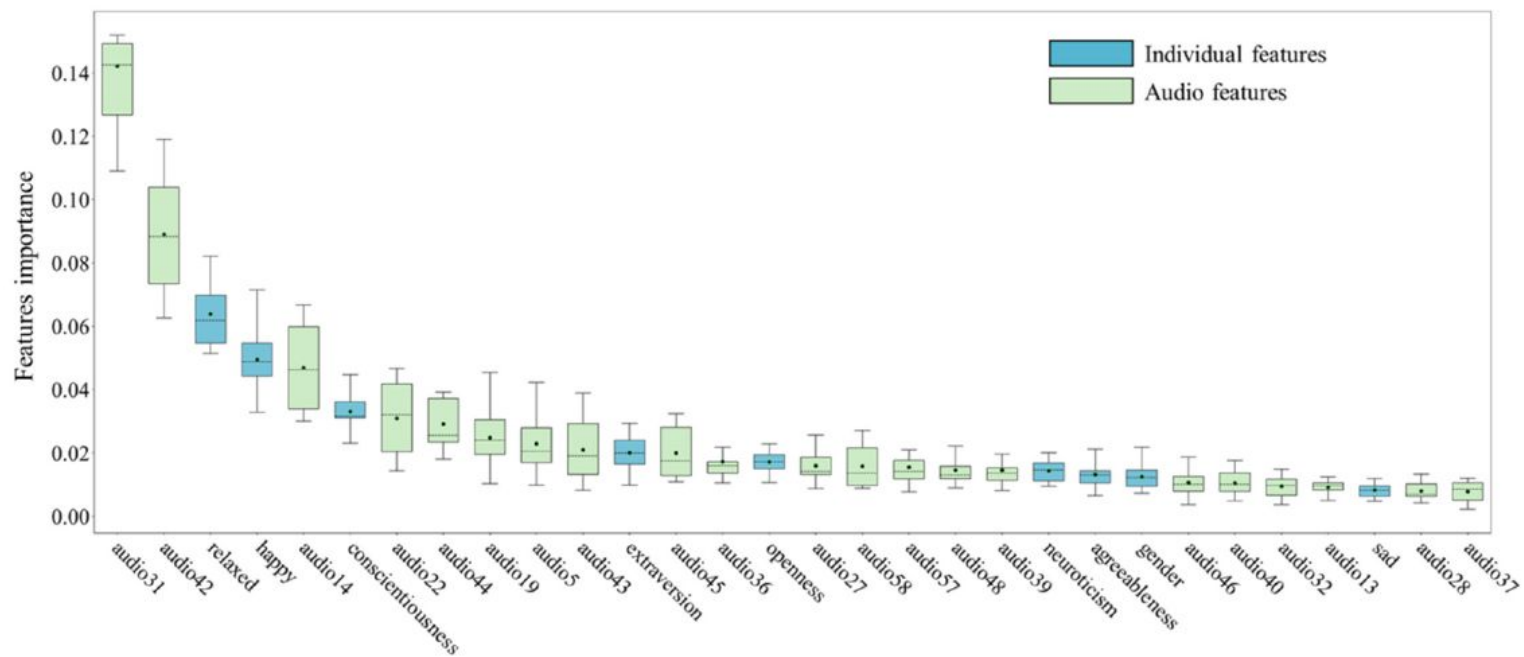
將性向測驗資料與悲傷程度資料正規化。

使用MLR及RFR作為主要驗算法，並比較了三種不同資料輸入方法。

Parameters	Inputs		
	AF	IF	CF
<i>n_estimators</i>	17	39	99
<i>max_depth</i>	38	37	18
<i>min_samples_leaf</i>	28	47	23
<i>min_samples_split</i>	42	17	5
<i>max_features</i>	0.6	0.5	0.6

AF indicate audio features; IF indicate individual features; and CF indicate combined features.

研究結果





使用深度學習的音樂轉 錄建模和作曲



Introduction

人工神經網絡在音樂建模、作曲和聲音合成中的應用. 但新的東西是前所未有的資源計算能力到數據, 卓越的培訓方法, 以開放和可重複的研究。這種可訪問性是“深度學習”方法在機器學習的許多應用中遠遠超出最先進結果的一個主要原因, 例如, 圖像內容分析、語音處理和識別、文本翻譯, 以及更具創造性的藝術風格轉移, 以及 Google 的 Deep Dream。只要應用領域“數據豐富”, 深度學習方法就會做出重大貢獻。深度學習現在正應用於音樂數據, 從分析和建模錄音內容到生成新音樂

LSTM, RNN

深度神經網絡是一種在其輸入層和輸出層之間具有多個隱藏單元層的網絡。本質上，神經網絡通過一系列級聯非線性操作來轉換輸入。

Recurrent neural network(RNN) 是任何具有從至少一個單元的輸出到位於比自身更淺層的另一個單元的輸入的直接連接的神經網絡

將音樂描述為一系列符號使RNN 立即適用於對其進行建模

構建和測試的RNN由具有19個單元的輸入層、具有8-15個單元的單個隱藏層和具有15個單元的輸出層組成。每個輸入和輸出層中的一個單元是“音符開始”狀態；其他14個單位代表音高，從D4到C6各一個。其他四個輸入單元識別特定的單音訓練旋律，其中有四個，每個4個音符長。它劃分時間，使得模型的每個時間步長代表一個八分音符的持續時間。

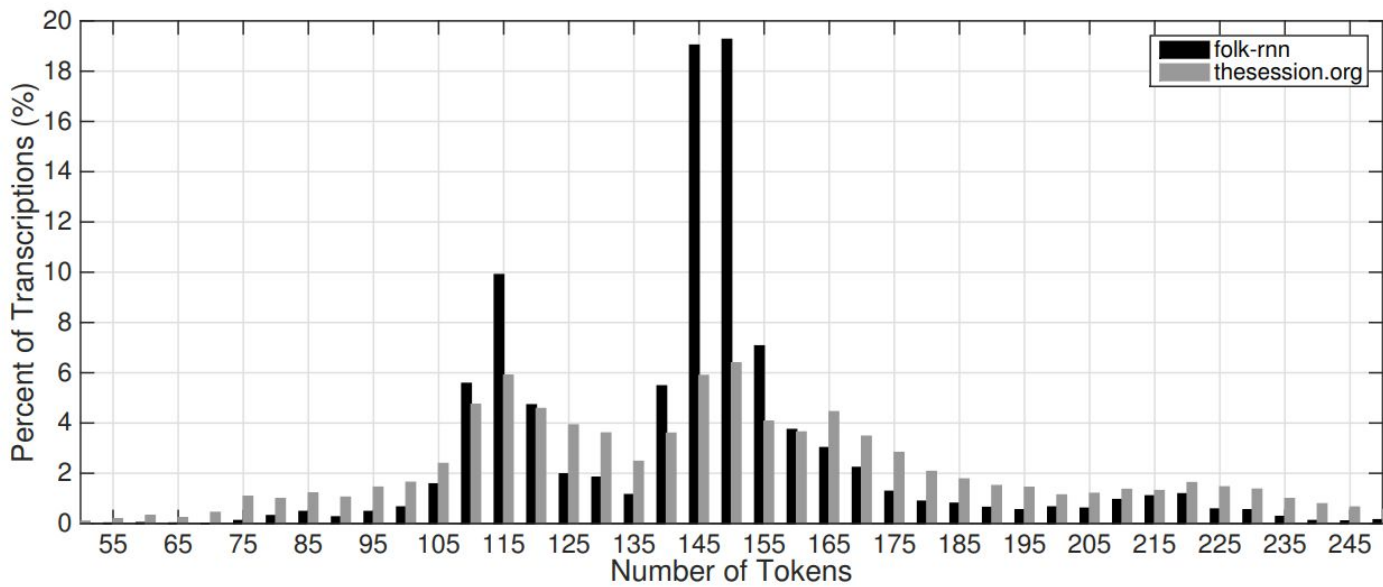
設計LSTM networks

所有的 LSTM 網絡都具有相同的架構，但在不同的詞彙表上運行，並且經過不同的訓練。我們構建的一種，我們稱之為 char-rnn，對單個字符的詞彙表進行操作，並在連續的文本文件上進行訓練。

```
3038,3038,"A Cup Of Tea","reel","4/4","Amixolydian","|:eA (3AAA g2 fg |  
eA (3AAA BGGf|eA (3AAA g2 fg|1afge d2 gf:|2afge d2 cd||  
|:eaag efgf|eaag edBd|eaag efge|afge dgfg:|","2003-08-28 21:31:44","dafydd"  
3038,21045,"A Cup Of Tea","reel","4/4","Adorian","eAAa ~g2fg|eA~A2 BGBd|  
eA~A2 ~g2fg|1af (3gfe dG~G2:|2af (3gfe d2~cd||eaag efgf|  
eaag ed (3Bcd|eaag efgb|af (3gfe d2~cd:|","2013-02-24 13:45:39",  
"sebastian the megafrog"
```

結構

我們構建的每個 LSTM 網絡都有三個隱藏層，每個隱藏層有 512 個 LSTM 塊，輸入和輸出單元的數量等於其詞彙表中的字符或標記的數量。我們以本地方式對我們的轉錄進行編碼，例如 in，其中詞彙表中的每個元素都映射到一個輸入和輸出單元。每個網絡的輸出是其詞彙表的概率分佈。



我們從 char-rnn 模型生成了 72,376 個曲調轉錄，並自動合成了其中的 35,809 個。我們使用這些結果創建了“無盡的傳統音樂會議”，

16-measure musical score in D major (two sharps) and 4/4 time. The score is written on a single staff with a treble clef. The melody consists of eighth and quarter notes, often beamed together. Fingerings (1-5) and articulation (accents, slurs, and breath marks) are indicated throughout. Chord symbols (I, V, IV, ii) are placed below the staff at various points. A first/second ending bracket is present over measures 8 and 9.

Measures and Chords:

- Measure 1: Chord I
- Measure 2: Chord V
- Measure 3: Chord I
- Measure 4: Chord V
- Measure 5: Chord I
- Measure 6: Chord V
- Measure 7: Chord IV
- Measure 8: Chord V
- Measure 9: Chord I
- Measure 10: Chord V
- Measure 11: Chord I
- Measure 12: Chord ii
- Measure 13: Chord V
- Measure 14: Chord I
- Measure 15: Chord V
- Measure 16: Chord I

T: Mal's Copporim, The

M: 4/4

L: 1/8

K: Dmaj

|: a>g | f2 f>e d2 d>B | A>BA<F A2 d>e | f2 d>f e<ac>d | e>dc>B Agfe |
f2 f>e d2 d>B | A2 A>G F2 F2 | G2 B>A d2 c>d |[1 e>dc>A d2 :|[2 e2 d2 d2 ||
|: f<g | a>Ag>A f>Ae>A | d>gd>B d2 g>A | f>Af>e d>ed>c | e>ed>c (3Bcd (3efg |
a2 a>g f2 e2 | d2 A>d f2 f>g | a2 g>f e2 f>g | a2 A2 D2 ||