Research Article

A Deep Learning-Enabled Composition System Based on Piano Score Recognition

Rui Shang

College of Music, Guangdong Polytechnic Normal University, Guangzhou, Guangdong 510000, China

Correspondence should be addressed to Rui Shang; shangrui@gpnu.edu.cn

Received 2 May 2022; Revised 5 June 2022; Accepted 13 June 2022; Published 5 July 2022

Academic Editor: Abid Yahya

Copyright © 2022 Rui Shang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Piano is used for music and comprises a stringed keyboard instrument wherein the strings are tapped by softer-coated wooden hammers. The score providing music for the piano, often a compressed transcription of orchestral music, is referred to as piano score. Presently, the Internet is overflowing with music score resources. Having so many music score resources available, professional learners and amateur music lovers are unable to identify and obtain music score information that matches their needs and wasting valuable time. Due to the rapid development of deep learning algorithms, some individuals utilize these algorithms to detect piano scores and construct composition systems, reducing the need of traditional machine learning algorithms on manual design and music knowledge guidelines. This paper uses the deep learning algorithm to construct piano score recognition framework based on K-Nearest Neighbor (KNN) algorithm and formulates the recognition system into multinote that significantly improves the recognition rate for the system. The self-attention mechanism is then introduced in order to build a composition system based on deep learning in which composition training and processes are described. Finally, a comparative experiment is conducted to evaluate the recognition accuracy for the KNN-based piano score recognition system. The results show that highest recognition accuracy of this system is 67.5%. The effect of composition system is evaluated based on prediction accuracy of notes. Three experiments are conducted to train the composition notes. As a result, the prediction accuracy of experiments 1, 2, and 3 is 89.2%, 91.8%, and 92.7%, respectively, indicating that the system has a high prediction accuracy and a perfect composition effect.

1. Introduction

The piano resembles a string keyboard equipment for which strings have been hammered by hardwood hammers covered with a softer substance [1]. This is played on a keyboard that is series of keys, in which the user presses or strikes with both hands’ fingers as well as thumbs to trigger the hammers to beat the strings. A piano score is a musical score that condenses the numerous instruments sections onto two staffs. At present, most experts in the field of music score recognition focuses on the recognition of music types and music emotional classification and so on [2]. Some of the experts are working on the recognition of piano music scores. As there is a large amount of piano music score data stored on the network in the world of big data, it is difficult for individuals to correctly identify and retrieve piano music score. At the same time, the composing process is also complex; therefore, music lovers cannot obtain a high number of network-required composition resources that makes composing more difficult. For this reason, this paper uses deep learning algorithm in the context of piano score recognition to solve this problem [3].

In order to address this issue, this research develops a piano music score identification framework based on the KNN deep learning algorithm and employs this technique to reliably identify piano music score data in the network [4]. Simultaneously, a composing system based on deep learning algorithms is being developed, and users can generate a complete music work, depending on their demands by simply entering data into the system [5].
The innovations in this research process are as follows:

(1) In this paper, the KNN deep learning algorithm is used to build a piano music score recognition structure, which is then used to train and preprocess piano music score data, extract piano music score for recognition, and improve the accuracy and stability of music score recognition.

(2) Build a composing model by introducing attention mechanism based on deep recognition algorithm, then introduce the training and deep learning processes of composing. Finally, improve the accuracy rate of composing prediction.

The structure of the remaining paper is as follows. Section 2 is composed of the related work, Section 3 is the piano score recognition based on deep learning, Section 4 is the composing system based on deep learning algorithms, Section 5 is the analysis of composition system of deep learning algorithms based on piano music score recognition, and finally, the paper is concluded in Section 6.

2. Related Work

The theory and technology of in-depth learning have advanced dramatically in recent years. Music recognition and composition systems have also advanced fast as a result in the development of in-depth learning algorithms and are now a hot research topic for both domestic and international researchers. In [6], the authors used the transformer Artificial Neural Network model in the field of piano music generation. The main components of their model are attention units. Compared with traditional recurrent neural network (RNN) algorithms, such as Long Short-Term Memory (LSTM), this model demands high modelling skills and can learn scale relationships on the sequence completely. The model can generate extended piano polyphonic music with a strong theme. In [7], the authors used a training RNN model in piano performance; this model not only generates music, but it can also learn dynamic music performance. Na et al. investigate the effectiveness of deep learning algorithms in automatic music generation, but having less literature on the interactivity for the generation systems of neural music because interpretability of neural networks is poor [8]. Dua et al. introduced the generation rhythm network based on LSTM; it is made up of a forward feedback network and an LSTM, where the LSTM is in charge of learning drums while the forward feedback network is in charge of mastering rhythm data. Fusing the outputs of the two networks yields the result [9]. Sun et al. have introduced a music generation system based on LSTM that can effectively control and adjust music styles as well as recognize 23 different styles of music [10]. Xia et al. formulate convolution neural network music target detection tasks for different regions. A complete image block is divided into small pieces by using sliding window and context correlation. Each block can only have one mean map. After running faster Region-Based Convolutional Neural Network (R-CNN) detection model, all symbol categories and location data in clipped image are obtained [11]. Jiang et al. give a new method to deal with music target detection known as Deep Water Detector, which trains convolution neural network, customizes energy function to accomplish music detection task, and uses custom energy function to split music score semantics in watershed transformation. Huang et al. use sequence-to-sequence models and convolution neural networks to identify printed monophonic music scores [12]. Liu et al. use convolutional neural network (CNN) algorithm to extract features from printed music scores and use Connectionist Temporal Classification (CTC) loss function to handle alignment problems between real data and music score data [13]. Yu et al. use a convolution neural network to extract the music score image feature data from handwritten music scores and then use a two-way LSTM model to accurately identify handwritten music scores and develop a handwritten music score recognition benchmark [14].

3. Piano Score Recognition Based on Deep Learning

The piano score recognition based on deep learning section is divided into the following sections.

3.1. Deep Learning Concepts. Deep learning is a technological method based on deep neural networks that can simulate how the human brain collects deep-level properties of samples to precisely understand sample distribution, and its impact is superior to typical machine learning algorithms [15]. Neural networks are defined by simulating the patterns of information transmission and processing among neurons in the human brain. The two successful applications of deep learning are natural language processing and computer vision. From the above analysis, it is found that the deep learning in neural network and machine learning are directly related while neural network is the primary technical measure and method. Starting with the challenges encountered by each specialization and integrating them with the problem-solving method, a neural network is chosen and modified to attain higher processing efficiency and better results.

3.2. Piano Score Recognition Based on KNN Algorithm. KNN is the simplest deep learning classification method. This paper presents a piano music score recognition based on KNN algorithm. In the recognition process, new difficulty features and existing features are employed as the feature space, and deep learning concept is used to optimize the problem based on the maximum interval that needs to be calculated. Prior information from the training data is used to generate the projection matrix and the original features are projected into space with high class discrimination [16]. When completed, the piano score is recognized using the classification principle of the KNN algorithm. Figure 1 shows the structure framework of piano music score recognition based on KNN algorithm.

The digital music score data is judged based on the deep learning theory by analyzing the identification process depicted in Figure 1 and the projection matrix M with a high degree of similarity for piano difficulty is accurately identified. The projection matrix is used to project the music
score into the space with high degree of category discrimination. After completion, the piano music score is recognized by using the KNN algorithm classification principle. This algorithm is adaptable to different music score categories and has improved recognition accuracy and stability. It provides reliable piano score recognition data for music learning and piano teaching [17].

In order to obtain the best feature projection matrix for the algorithm, the $L \in \mathbb{R}^{n \times n}$ feature transformation matrix can be used for feature projection:

$$x'_i = Lx_i.$$ (1)

The $L$ element in the linear combination transformation matrix forms a new projection eigenvector $x'_i$ with the original eigenvector of $x_i$. Calculate the distance measure in $d_L$ projected space from the following formula:

$$d_L(x_i, x_j) = \|L(x_i - x_j)\|_2 = \sqrt{[L(x_i - x_j)]^T [L(x_i - x_j)]}.$$ (2)

In equation (2), $x_i$ and $x_j$ both are eigenvectors. Selecting the square value of the distance to be expressed as a matrix ensures that the distance value is a positive number. By using equation (3), calculate the new $D_M$ measurement distance:

$$D_M = D_L^2 = [L(x_i - x_j)]^T [L(x_i - x_j)]$$

$$= (x_i - x_j)^T L^T L (x_i - x_j)$$

$$= (x_i - x_j)^T M L (x_i - x_j).$$ (3)

In equation (3), $M$ is a matrix that represents a semi-positive definite symmetric matrix, $M = L^T L$, while the use of this matrix is to project features to other spaces and anticipating a greater class differentiation after feature space projection. In short, the notes in the piano music score with the highest similarity after projection are in the same label, and the music scores with different labels are highly dissimilar. Projection will bring music with higher similarity closer together, while it will expand the gap between music with lower similarity, at least beyond that of the same type of data. Following these principles, the method of calculating the $M$-projection matrix is described as the objective function shown as follows:

$$\min_{M \succeq 0} \sum_{(i,j) \in S} D_M(x_i, x_j),$$

s.t. $\sum_{(i,j) \in S} [D_M(x_i, x_j) - D_M(x_i, x_j)] \geq 1.$ (4)
Select the $M$ matrix with the smallest value, $S$ is a dataset consisting of two numeric scores for the same grade, and $R$ is a set that is formed by adding a new grade to the numerical scores of different grades to produce three sets. The distance between labels from different categories surpasses the distance between labels of the same kind by one unit. This operation can maximize the class discrimination. The optimization process is shown in Figure 2. The purpose of the optimization is to shorten the distance between the same types of music scores and to increase the distance between different types of music scores. The optimization principle is similar to Large-Margin Nearest Neighbor (LMNN) algorithm.

To identify new piano music score data, it is needed to preprocess the characteristics and extract the same type of features and then learn the label data for different levels of piano music score to get $M$-matrix, and after that, use equation (3) to calculate the distance between each music score in the training set. The results will show some similarity.

3.3. Multinote Recognition System. This research focuses on two forms of recognition based on the differences between piano music score recognition tasks: multinote and single-note recognition. Recognition accuracy of piano single note is up to 98%, which meets the practical requirements. However, when recognizing piano multinoites, the recognition rate is less than 70%. This paper establishes a multinote recognition system to improve the recognition rate.

A semantic model and a phoneme model are produced when recognizing continuous speech with a huge amount of data. The phoneme model is an acoustic model and the smallest unit to build a Hidden Markov Model (HMM). Similar to voice recognition systems, semantic models can be used to do statistical analyses of the connections between distinct factors or to compute the probability of showing the next component based on known current factors. The multinote recognition system based on HMM combines the internote model with the multinote model. The multinote model serves as the foundation for the fundamental recognition concept. The degree of match between the test audio and the multinote model is calculated, and $P(O|M_i)$ probability value is calculated. If the number of HMM multinote models in a model library is $N$, the calculated probability of $N$ notes is expressed by $P(O|M_i)$. The difference between the maximum value in probability and other probabilities is calculated using the following formula:

$$\delta_i = P_N - P_i, \quad i = 1, 2, \ldots, N - 1.$$  \hspace{1cm} (5)

If the number of $\delta$ is less than that in $\{\delta_i\}$, then the threshold value can be regarded as the recognition result of the input frame obtained by the model corresponding to $P_N$. If the memory of $\{\delta_i\}$ is less than $\delta$ threshold value, then the statistical model between multinotes is investigated to clarify this multinote name, and the ideal model is generated by utilizing the following equation:

$$M = \arg \max_{i \in S} (P(O|M_i)) + P(w_t|w_{t-1}w_{t-2}).$$ \hspace{1cm} (6)

By using equation (6), the entire set of elements for the threshold is represented by $S$.

Here, the HMM-based multinote recognition process is simplified as $\max\{P(O|M_i)\}$, where $i$ represents the first arranged multinote model.

3.4. Piano Multinote Recognition Rate. Based on the preceding method, the most important aspect of obtaining the result of multinote recognition system identification is to accurately evaluate the system recognition effect, in which the selection of multinote is evaluated by the right rate. This study primarily compares and tests the audio recognition results, the initial time, the end time, and the multinote name between the reference labels. The accuracy of the piano multinote recognition system is assessed upon it. The following equation is given for calculation:

$$P_e = \frac{N - N_F - N_D}{N}.$$ \hspace{1cm} (7)

In equation (7), $N$ denotes the total number of multinotes in the reference label, $N_D$ represents the number of multinotes greater than the initial and stop time intervals, and $N_F$ denotes the number of multinote recognition errors during initial and stop time intervals.

4. Composing System Based on Deep Learning Algorithms

This section is divided into the following three sections.

4.1. Build Composing Models. Context information can affect the process of music generation to some extent. This problem cannot be effectively handled only through the circular neural network structure. Therefore, the self-attention mechanism needs to be introduced. With the implementation of this method, greater emphasis should be placed on the essential information in the context.
Secondary information should be ignored to make the generated notes more accurate. The self-attention mechanism can also assign probability weights to corresponding contextual note features, deeply mine the dependent features in the note sequence, and achieve accurate expression of the note sequence, which makes the prediction and generation of note sequence more accurate as well as ensures the validity of music generation [18]. Figure 3 shows the structure of composing model built in this paper.

The fundamental framework of the composing model is depicted in Figure 3. The composing model is divided into three sections after evaluating the composing structure, such as hidden layer, input layer, and output layer. Synchronized many-to-many mode is used in input and output mode while the length of note sequences in the input and output are identical, and they are quite long.

4.2. Composition Training Process Based on Deep Learning. The synchronous multiple-input, multiple-output mode in the loop neural network is used in the composing model. The n-length musical sequence is used as the input data in this model and the length of the output musical sequence is also n [19]. Therefore, great attention should be paid to the sequence of melody notes in the diced composition. It is assumed that Notes = \{note_1, note_2, ..., note_i\} is the melody note sequence, while the following X-matrix is obtained by slicing the note sequence.

\[
X = \begin{bmatrix}
\text{note}_1 & \text{note}_2 & \ldots & \text{note}_n \\
\text{note}_{n+1} & \text{note}_{n+2} & \ldots & \text{note}_{2n} \\
\ldots & \ldots & \ldots \\
\text{note}_{i+n} & \text{note}_{i+n+1} & \ldots & \text{note}_{(i+1)n}
\end{bmatrix},
\]

(8)

The following Y note matrix for expected output is obtained:

\[
Y = \begin{bmatrix}
\text{note}_{n+1} & \text{note}_{n+2} & \ldots & \text{note}_{2n} \\
\text{note}_{2n+1} & \text{note}_{2n+2} & \ldots & \text{note}_{3n} \\
\ldots & \ldots & \ldots \\
\text{note}_{(i+1)n+1} & \text{note}_{(i+1)n+2} & \ldots & \text{note}_{(i+2)n}
\end{bmatrix}.
\]

(9)

Musical notes are represented by matrix that is the input data for the composing model. After slicing, selecting the sequence bits of the input note sequence predicts the output notes.

Input note sequence is as follows:

\[
X_i = [\text{note}_{i+n}, \text{note}_{i+2n}, \ldots, \text{note}_{i+2n}].
\]

(10)

Expected output note sequence is as follows:

\[
Y_i = [\text{note}_{(i+1)n}, \text{note}_{(i+2)n}, \ldots, \text{note}_{(i+2)n}].
\]

(11)

Composing model training rules: Input a music character sequence of a specified length into the composing model and forecast the output result of the following output music note sequence. The predicted output note sequence is then compared to the expected output note sequence in the training set, the cross-direct function is used to find the error between the expected and actual output results, and the back-propagation algorithm is used to update the learning parameters in the composing model. Many learning and training iterations occurs until the model remains convergent and the generated note prediction model becomes perfect [20].

4.3. Composing Process Based on Deep Learning. The composing model needs to be learned and trained several times to obtain a strong convergence prediction model. The composing process is based on this model and generates a specific length of note sequence. Music note prediction model inputs a music note sequence with length n. After first input, the note sequence can be randomly selected from the test set and entered into the music prediction model to output the next one with length n. The sequence of output note prediction is then utilized to forecast the following input value. After several iterations, it stops when generating...
the preset music length. A composing work will be created, and the composing model will produce a complete music according to the above composition. The length of the music character sequence is selected for input in the test set is \( n \), and the note prediction model generates a whole piece of music. This paper uses deep learning algorithm to build composing model, which can automatically generate music works. The following matrix is the input dataset required to form a music using the note prediction model in automatic composing model:

\[
X = \begin{bmatrix}
\text{note} & \text{note}_2 & \ldots & \text{note}_n \\
note_{pre1} & \text{note}_{pre2} & \ldots & \text{note}_{pre_n} \\
\vdots & \vdots & \ddots & \vdots \\
note_{pre_{i\times\alpha}} & \text{note}_{pre_{i\times\alpha+1}} & \ldots & \text{note}_{pre_{(n-1)\times\alpha}}
\end{bmatrix}.
\]

(12)

The following matrix represents the note sequence output by the note prediction model:

\[
Y = \begin{bmatrix}
\text{note}_{pre1} & \text{note}_{pre2} & \ldots & \text{note}_{pre_n} \\
\text{note}_{pre_{n+1}} & \text{note}_{pre_{n+2}} & \ldots & \text{note}_{pre_{2n}} \\
\vdots & \vdots & \ddots & \vdots \\
\text{note}_{pre_{(i\times\alpha+1)}} & \text{note}_{pre_{(i\times\alpha+2)}} & \ldots & \text{note}_{pre_{(i\times\alpha+\alpha)}}
\end{bmatrix}.
\]

(13)

The first row of data in the \( X \) matrix is to randomly select the sequence of notes in the test set when a musical work is generated, which corresponds to the output musical note sequence generated by the first row on matrix \( Y \). Then, the value of \( Y_1 \) is assigned to \( X_j \) as the input data for the next prediction. The resulting musical note sequence is calculated iteratively in this way, and finally a musical work note sequence is formed by splicing the sequence of notes generated in \( Y \) in each line. Figure 4 below shows the composing process of the composing model.

5. Analysis of Composition System of Deep Learning Algorithms Based on Piano Music Score Recognition

The analysis of composition system of deep learning algorithms based on piano music score recognition is divided into the following sections.

5.1. Piano Music Score Recognition Analysis. The experimental dataset must be determined in order to validate the piano music score recognition system based on the KNN algorithm. Piano music score data contains Musical Instrument Digital Interface (MIDI) format digital music score files with music score information such as beat, pitch, speed, chord, channel, and speed. Although piano teachers and music websites categorize piano scores into four levels, these levels are set when evaluating the scalability of the algorithm. The 400 MIDI scores on different websites are used to compose the four datasets of different levels; each level has 100 MIDI scores. NineS is the dataset of nine levels, while FourS is the dataset of four levels. The recognition accuracy of KNN-based algorithm and linear kernel function support vector machine (SVM) such that Lagrangian support vector machine (L-SVM) and polynomial kernel function SVM such that proximal support vector machine (P-SVM) algorithm datasets are verified by simulation experiments. The results are listed in Table 1.

According to Table 1, L-SVM, P-SVM, and KNN algorithms have different recognition results on NineS and FourS datasets, while the recognition accuracy of the three algorithms in FourS dataset is higher than that of NineS dataset. By comparing with three algorithms, the recognition accuracy of L-SVM algorithm on NineS dataset, P-SVM, and KNN-based algorithm is 63.7%, 62%, and 68.8%. The recognition accuracy of the three algorithms on the FourS dataset is 70.3%, 79.2%, and 85.6%, respectively. It is observed that the KNN-based algorithm has the highest recognition accuracy among the three algorithms.

To evaluate the system’s recognition impact on piano music score, 150 piano tunes were randomly picked as training sample data from 182 piano tunes that fit the criteria, and the remaining 32 were employed as system test data. The average length and number of multitones of a piano track were tested and trained; moreover, the result reveals that the average number of multicharacters is 719 and the average duration is 310 ms. The multitone created after training 150 piano recordings perfectly covers the multitone combination formed by 88 single notes on the piano, according to an analysis of the training model library. At the same time, Unconstrained Nonnegative Matrix Factorization (NMF), Constrained NMF, and Auditory Model are selected to compare and analyze the accuracy of the proposed KNN algorithm-based piano music score recognition system, which are shown in Figure 5.

Compared with other piano music score recognition accuracy rates, the recognition system used in this paper has an accuracy rate of more than 5%. Piano music score recognition based on the KNN method requires training samples in advance like contrast statistics, time-frequency domain, and other recognition techniques. However, this algorithm can accurately convert the multitone features to build a model, which is an important basis for multitone recognition. Analysis of Figure 5 shows that the recognition accuracy of Unconstrained NMF, Constrained NMF, and Auditory Model is 61.8%, 62.6%, and 59.8%, respectively, and the highest recognition accuracy of this system is 67.5%. The recognition accuracy meets the piano music score recognition requirements and can accurately identify piano music score data on the network.

5.2. Composition Effect Evaluation. This work develops a composing model based on an in-depth learning algorithm and objectively assesses the system composing impact by analyzing the composing model’s note prediction results. The accuracy of note prediction has been used as the assessment criterion in this case. Assume that \( N(i, j) \) is used to predict the number of notes, in which the actual note number \( i \) is \( j \). The accuracy represents the overall prediction accuracy of the notes. This value is derived as follows,
depending on the proportion of the precise number of notes in the overall number of notes:

\[
\text{accuracy} = \frac{\sum N(i,i)}{\sum \sum N(i,j)}
\]  \hspace{1cm} (14)

In this study, while validating, the highest probability of selecting a note number equal to the actual number is predicted properly, while the remainder are errors in the prediction. Experiments 1, 2, and 3 are used to evaluate the system’s composing abilities. Table 2 displays the final prediction accuracy values.

Experiment 3 has the greatest note prediction accuracy rate of 0.826 when compared to the others. Figure 6 depicts a curve that verifies the accuracy of vowel prediction during
According to the results of three experiments shown in Figure 6, the accuracy rate of experiment 3 is higher than that of experiments 1 and 2 during the training. The highest prediction accuracy rates of experiments 1, 2, and 3 are 89.2%, 91.8%, and 92.7%. It verifies that the composing model has a higher accuracy rate of prediction for notes and can meet people’s composing requirements.

6. Conclusion

With the rapid development of economy, people’s material life is becoming richer and some people begin to pursue spiritual satisfaction. Music art is an essential part of the spiritual world. People create a comfortable and tasteful living environment by connecting music and life together. Manual composing requires fundamental music theory knowledge, music format, debugging procedures, and so on, which is very professional for ordinary composition lovers. The piano is a musical instrument that consists of a strung keyboard instrument whose strings are touched by softer-coated wooden hammers. The piano score is a music score for the piano that is typically a reduced copy of orchestral music. The Internet contains a large number of piano music scores, and the people who want to improve their personal ability cannot find the required content due to a large number of piano music scores. In dealing with such issues, the deep learning algorithm is employed to build a piano score recognition system and a composition system based on the KNN method. After verification, the recognition accuracy of the piano score recognition system based on KNN algorithm is 67.5%. The prediction accuracy of notes is selected to analyze the effect of the composition system. After conducting three experiments to train the composition notes, the prediction accuracy of experiments 1, 2, and 3 is 89.2%, 91.8%, and 92.7%, indicating that the proposed system has a high prediction accuracy and an optimal composition effect.

Data Availability

All the data are included in this paper for publication.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References


