SUGISAKA MASANORI

Journal of Robotics, Networking and Artificial Life
Vol. 10(3); December (2023), pp. 282–285
ISSN (Online): 2352-6386; ISSN (Print): 2405-9021
https://alife-robotics.org/jrnal.html

Research Article

A Research on a System Using Deep Learning for Inferring Piano Performance

Yoshiki Hori, Eiji Hayashi
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan

ARTICLE INFO

Article History
Received 30 November 2023
Accepted 05 April 2024

Keywords
Automatic piano
Computer music
Deep learning

ABSTRACT
Achieving expressive and human-like automated piano performances has proven challenging. This study proposes a deep learning system to infer expressive nuances from musical scores, addressing the limitations of traditional rule-based approaches. By leveraging neural networks to learn the mapping between scores and expert performances, the system automates the inference process, improving accuracy while enhancing efficiency. This novel application of deep learning shows promise for advancing automated music performance and enabling more artistically expressive renditions. The insights gained could have broader implications for computer-aided musical interpretation and synthesis.

1. Introduction

Automated piano performance has been a longstanding challenge in the field of computer music. While previous systems (Fig. 1) have achieved precise control over keystroke mechanics, replicating the nuanced and expressive qualities of human performances remains an elusive goal. This study aims to develop a system capable of inferring the intricate musical nuances from score data, enabling natural and human-like piano renditions.

Fig. 1 Automatic piano playing device [1]

The advent of deep learning techniques has opened up new avenues for modeling complex patterns in data. By leveraging the power of neural networks, this research investigates the potential of deep learning algorithms to capture the subtle variations in timing, dynamics, and articulation that characterize expressive piano performances.

Unlike conventional rule-based or parametric approaches, deep learning offers a data-driven approach to learn the intricate relationships between musical scores and expressive performances directly from data. This novel application of deep learning to piano performance inference promises to bridge the gap between mechanical precision and artistic expression.

The proposed system not only advances the state of the art in automated piano performance but also holds broader implications for computer-aided music interpretation and synthesis. By enabling more natural and human-like renditions, this research paves the way for enhancing the artistic capabilities of automated music systems.

Corresponding author E-mail: hori.yoshiki311@mail.kyutech.jp, haya@mse.kyutech.ac.jp

Conventional systems could replicate expert performances by interpreting pitch and dynamics information encoded in MIDI data. However, a human-like expressive rendition could not be achieved from score data alone. To address this limitation, efforts have been made to develop systems that can generate expressive performances directly from musical scores. As an initial step, attempts were made to replicate the performances of renowned pianists.

However, previous approaches required manual inference of the expressive nuances for each individual note, a process that was exceedingly time-consuming and labor-intensive, even for short musical pieces containing thousands of notes. Therefore, in this study, we constructed a system that automates the inference process by leveraging deep learning techniques.

This novel approach, employing deep learning for automated inference, aims to overcome the limitations of conventional methods that relied on manual annotation. By enabling neural networks to directly learn the mapping between musical scores and expressive nuances, efficient and scalable inference can be achieved without the need for human intervention.

2. Inference system with Deep Learning

2.1. *Dataset*

In our study, we analyzed performance metrics from the renowned pianist Vladimir Davidovich Ashkenazy, captured using the MIDI protocol.

2.2. *Performance data*

Given the inherent complexity of raw performance data for straightforward use in machine learning, this study segmented the data into distinct categories: performance details and score particulars. By establishing specific parameters, we were able to refine and structure the data, rendering it more accessible for machine learning algorithms. This methodology not only clarifies and systematizes the data but also optimizes it for algorithmic learning, thereby improving the applicability and efficiency of machine learning techniques in analyzing piano performance interpretation.

In Sections 2.3 and 2.4, the research details the parameters set for both performance information and score information.

2.3. *Performance information*

In this research, the performance information encompasses the expressive qualities demonstrated by the pianist in the collected data. For the purpose of analysis, we established four specific parameters to describe this performance information. In this research, we established four parameters to characterize the performance information. The details of these performance parameters are presented in Table 1.

Table 1 The performance information format

Parameter	Units	About
Velo	None	Sound intensity
Gate	ms	Length of note
Step	ms	Interval between the next note
Time	ms	Time of sound

In this study, Time is not an object of deduction as it can be derived from the Gate and Step values.

2.4. *Score information*

Score information consists of performance data featuring musical notations, including notes and symbols. In this study, we defined five parameters based on the score information. These parameters are detailed in Table 2.

Table 2 The score information format

Parameter	Units	About
Key	None	Sound height
Bar	None	Bar number
Dyn	None	Dynamics mark
Tgate	ms	Length of note on the score
Tstep	ms	Interval between the next note in the score

2.5. *Neural network configuration*

Existing inference systems have predominantly relied on traditional methods, with deep learning architectures remaining largely unexplored. Nevertheless, when creating a predictive model with deep learning, identifying the most effective network architecture for interpreting performance information is crucial. Consequently, we opted to develop and test a system

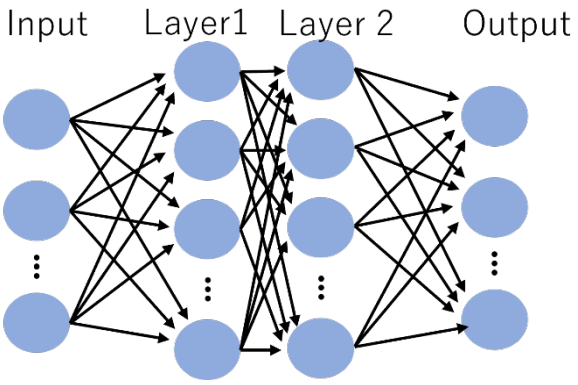


Fig. 2 System overview

featuring a straightforward network design. Fig. 2 illustrates the schematic of the network we constructed.

3. Inference Experiment and Results

3.1. *Data splitting*

In this study, the performance data was segmented into sets for training, validation, and testing. K-fold cross-validation, a method used to assess the model's ability to generalize, involves splitting the data into K segments. Each segment serves as validation data in turn, while the remaining K-1 segments are used for training. This process ensures each data segment is used for validation once. Fig. 3 illustrates the concept of K-fold cross-validation. For the testing phase, "Prelude Op.28-7" by Fryderyk Franciszek Chopin was selected as the test data.

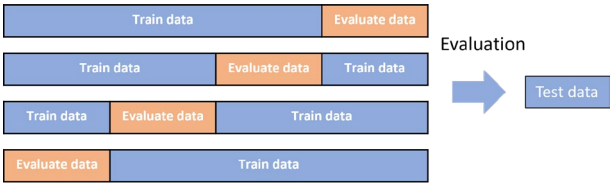


Fig. 3 Overview of K-fold Cross-Validation

3,2. *Validated network structure*

In this experiment, the number of neurons in the first and second intermediate layers was adjusted based on the patterns outlined in Table 3.

Table 3 Combination of neurons in intermediate layers 1 and 2

Pattern number	Layer 1	Layer 2
1	50	50
2	50	100
3	100	100
4	100	50

3.3. *Results*

The inference system generated graphs for each of Velo, Gate, and Step, comparing the deduced performance information with that of the pianist's actual performance. Fig. 4, Fig. 5 and Fig. 6 display the graphs for the patterns exhibiting the strongest correlation coefficients. Table 4 details the correlation coefficients between the performance information predicted by the system and the actual performance data from the pianist.

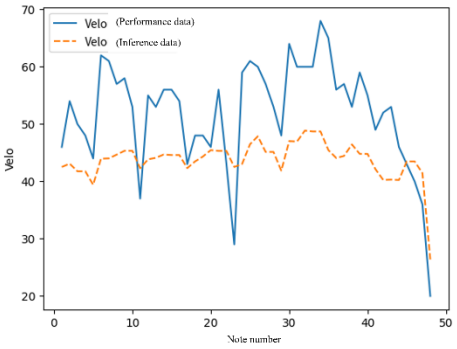


Fig. 4 Pattern 2 Velo

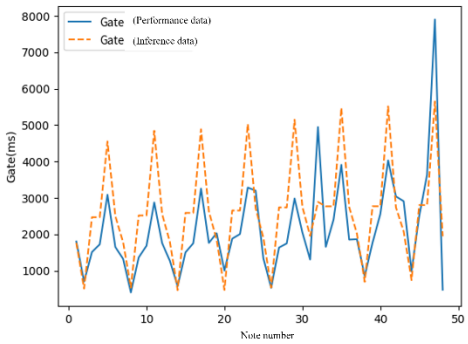


Fig. 5 Pattern 3 Gate

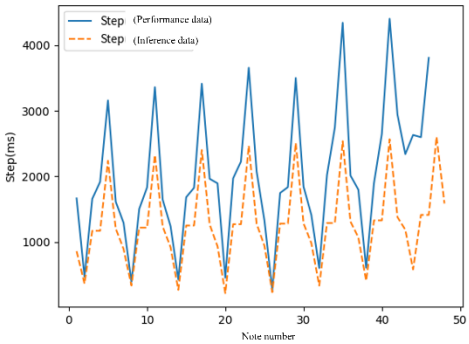


Fig. 6 Pattern 3 Step

Table 4 Performance information's correlation coefficient

Pattern number	Layer 1	Layer 2	Velo	Gate	Step
1	50	50	0.763	0.720	0.876
2	50	100	0.776	0.741	0.880
3	100	100	0.770	0.743	0.885
4	100	50	0.763	0.691	0.832

4. Consideration

Table 4 indicates that the correlation coefficient for Velo remains largely unchanged across all patterns. Despite Velo's high correlation coefficient, the graph waveforms display dissimilarity, as observed in Fig. 4. This discrepancy arises from the model interpreting points with significant value shifts as anomalies, leading to reduced variability in predicted values.

In contrast, the correlation coefficient for Step surpasses those of Velo and Gate, attributable to the Step waveform's more consistent variations, facilitating easier learning, as depicted in Fig. 6.

Moreover, an examination of the correlation coefficients for patterns 2 and 3 reveals high values for Velo, Step, and Gate alike. This implies that a network configuration incorporating a greater number of neurons might be more effective in capturing the nuances of these parameters. This points towards the potential benefits of employing a more intricate network structure to improve the fidelity of performance information inference, ultimately leading to more nuanced and precise modeling of musical expression.

5. Conclusion

This study explored the application of deep learning techniques for inferring expressive performance information from musical scores. We constructed and evaluated an inference system with a simple neural network architecture, focusing on identifying the optimal configuration for accurately capturing the nuances in piano performances.

The experimental results demonstrated that a network structure with a larger number of neurons in the layers closer to the output layer exhibited superior performance in inferring the intricate expressive details. This finding suggests that increasing the representational capacity of the deeper layers can effectively model the complex relationships between score data and expressive renditions.

While the proposed system represents a promising step forward, further enhancements are necessary to achieve more accurate and robust inference. Future work will investigate the incorporation of additional neurons, layers, and gating mechanisms that take into account previous inputs. By iteratively refining the network architecture and leveraging more advanced deep learning techniques, we aim to develop a comprehensive system capable of generating truly human-like and artistically expressive piano performances from score data.

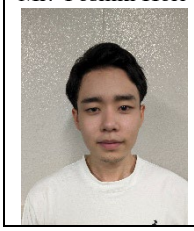
This line of research not only advances the state of the art in automated piano performance but also holds broader implications for computer-aided music interpretation and synthesis, potentially enabling a wide range of applications that bridge the gap between mechanical precision and artistic expression.

References

1. E. Hayashi, M. Yamane, H. Mori, Development of a moving coil actuator for an automatic piano, Int. J. Japan Soc. Prec. Eng. 28 (1994), 164–169.

Authors Introduction

Mr. Yoshiki Hori



He received bachelor degree in Engineering in 2023 from mechanical system engineering, Kyushu Institute of Technology in Japan. He is currently a Master student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Dr. Eiji Hayashi

He is a professor in the Department of Intelligent and Control Systems at Kyushu Institute of Technology. He received the Ph.D. (Dr. Eng.) degree from Waseda University in 1996. His research interests include Intelligent mechanics, Mechanical systems and Perceptual information processing. He is a member of The Institute of Electrical and Electronics Engineers (IEEE) and The Japan Society of Mechanical Engineers (JSME).