A Research on Performance Information Editing Support System for Automatic Piano

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

Eiji Hayashi
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan
Email: hori.yoshiki311@mail.kyutech.jp, haya@mse.kyutech.ac.jp

Abstract

In 1996, Hayashi et al. from our laboratory developed an automatic piano playing device that requires data with nuances for each note to perform in a human-like piano playing. However, this device lacks the function to infer such nuanced data. Therefore, prior research focused on developing a system to infer data with nuances for each note. Yet, this system required manual inference, consuming a significant amount of time. In this study, we have constructed a system capable of automatically performing inference using deep learning. This system not only improves the accuracy of piano playing inference but also contributes to the efficiency of the inference process.

Keywords: Automatic Piano, Computer Music, Deep Learning

1. Introduction

The automatic piano playing device (Fig. 1) developed by Hayashi and others, equips a grand piano's keyboard with striking mechanisms and its pedals with driving mechanisms, allowing for precise keystrokes and pedal operations through computer control. During its development, the behavior of the piano action was analyzed using a unique mechanical model, examining its dynamic properties. Additionally, to create optimal drive waveforms according to each key, the results of the key-wise behavior analysis were stored in a database for stable sound reproduction. The waveforms are generated by referencing the database in accordance with the music data. This technology enables stable performance, even replicating delicate and rapid repetitive strikes that are challenging for professional pianists.

Fig. 1 Automatic piano playing device [1]

The device plays as instructed by inputting MIDI (Musical Instrument Digital Interface) standard data, a common standard for transferring and sharing performance data among electronic instruments, containing information like pitch and volume. By recording a professional pianist's performance as MIDI data and inputting it into the device, it's possible to replicate their performance. However, due to the system's design, inputting only the score data does not result in a human-like performance. To address this issue, efforts are being made to develop a system that generates human-like performances from score data. As a first step, efforts are being made to replicate the performances of world-renowned pianists.

Previous research developed a system that infers data with nuances for each note, but this system required manual inference. Since even short musical scores contain over a thousand notes, the inference process was time-consuming and labor-intensive. Therefore, in this study, we constructed a system that automatically performs inference using deep learning.

2. Inference system using Deep Learning

2.1. *Dataset*

In this research, we used performance data of the pianist Vladimir Davidovich Ashkenazy, which were recorded in accordance with the MIDI standard.

2.2. *Performance data*

In performance data, the raw state is often too complex for direct machine learning application. Therefore, this research involved separating the performance data into performance information and score information and setting parameters to make it manageable for machine learning. This process essentially simplifies and organizes the data, making it suitable for the algorithms to process and learn from, thereby enhancing the effectiveness of the machine learning application in the context of 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*

Performance information includes data that encapsulates the pianist's expression within the performance data. In this study, four parameters were set for the performance information. These parameters of the performance information are shown in Table 1.

Table 1 The format of the performance information

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 research, Time is not a target of inference because it can be calculated from the Gate and Step values.

2.4. *Score information*

Score information is performance data that includes musical notation such as notes and musical symbols. In this research, five parameters were set from the score information. Table 2 shows the parameters of the score information.

Table 2 The format of the score information

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*

The inference systems developed to date have not been built using deep learning. However, in developing an inference system using deep learning, it is necessary to find the optimal network structure for inference of performance information. For this reason, we decided to construct and verify a system with a simple network structure. Fig. 2 shows a schematic of the constructed network.

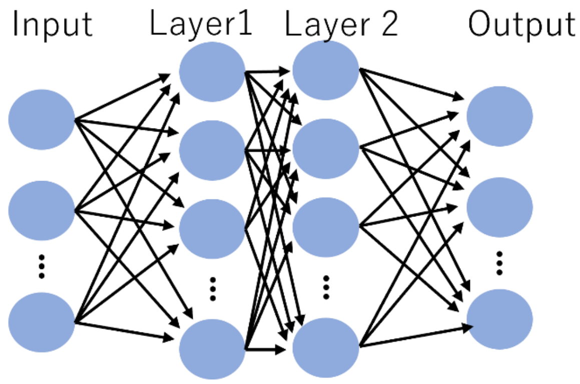

Fig. 2 Network overview

3. Inference Experiment and Results

3.1. *Data splitting*

In this experiment, the performance data was divided into training data, validation data, and test data. K-fold cross-validation is a statistical method for evaluating generalization performance in which the data is partitioned into K pieces, one of which is the validation data, and the remaining K-1 pieces are used as training data and evaluated with the test data. This is done by dividing the data into K pieces, one of which is the validation data. Fig. 3 shows a schematic of K-fold cross-validation. In this experiment, Prelude Op.28-7 by Fryderyk Francszek Chopin was used as the test data.

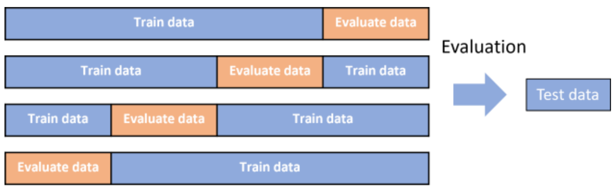

Fig. 3 K-fold Cross-Validation overview

3.2. *Validated network structure*

In this experiment, the layers of neurons in intermediate layer 1 and intermediate layer 2 were varied according to the patterns shown 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 performance information inferred by the inference system and the pianist's performance information were output as graphs for each Velo, Gate, and Step. The graphs of the patterns with the highest correlation coefficients are shown in Fig. 4, Fig. 5 and Fig. 6. The correlation coefficients between the inferred performance information and the pianist's performance information are shown in Table 4.

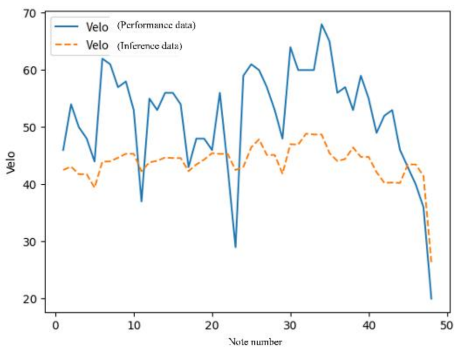

Fig. 4 Velo of pattern 2

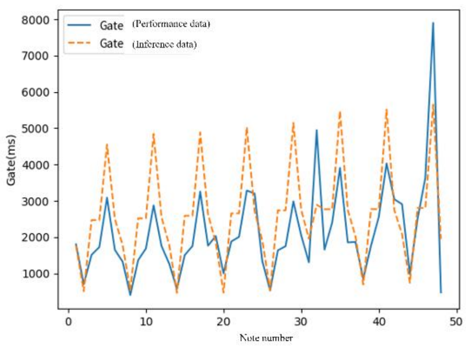

Fig. 5 Gate of pattern 3

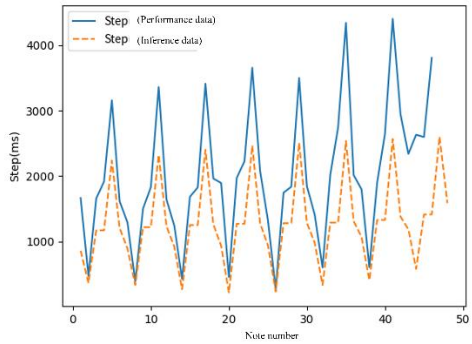

Fig. 6 Step of pattern 3

Table 4 Correlation coefficient of performance information

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 shows that Velo's correlation coefficient does not change significantly for any of the patterns, and although Velo's correlation coefficient is high, the waveforms of the graphs are not similar, as can be seen in Fig. 4. This is because the points with large value fluctuations were learned as outliers, and the fluctuations in the inferred values became smaller.

The correlation coefficient of Step is larger than that of Velo and Gate. This is because the variation of the Step waveform is more regular than the other values, making it easier to learn, according to Fig. 6.

Furthermore, looking at the correlation coefficients for patterns 2 and 3, we see that the correlation coefficients for Velo, Step, and Gate are all large. This suggests that a network structure with a large number of neurons in the layers close to the output layer is suitable for inference.

5. Conclusion

In this study, we developed an inference system with a simple network structure using deep learning. Using the developed system, we verified the optimal network structure for inference of performance information. The results showed that a network structure with a large number of neurons in the layers close to the output layer is suitable for inference.

The future goal is to develop a more accurate inference system. To this end, we will search for the optimal network structure by testing a system with an additional number of neurons, layers, and a gating mechanism that takes previous inputs into account.

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.

Prof. 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).