Research on performance information editing support system for automatic piano
- Development of a network model for improved dynamics accuracy-

Taiyo Goto
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan

Yoshiki Hori, Eiji Hayashi
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan
Email: gotou.taiyou711@mail.kyutech.jp, haya@mse.kyutech.ac.jp

Abstract

The automatic piano player, which was previously developed in this laboratory, is attached to the keys and pedals of a grand piano, and enables accurate keystrokes and pedal operation with appropriate control from a computer. To control the device, music data is required, but if music score data is simply input into the device, the performance will be flat, and will not sound like a human being, which is the goal. This is because pianists play with their own intonation when they play. Previous research has developed a system that uses deep learning to predict performance information, but the accuracy of predicting sound volume (Velo) was not good. This research aims to enhance the accuracy of Velo in automatic piano performance. A new deep learning system combining two networks was developed to address limitations in existing methods.

Keywords: Deep Learning, Automatic Piano, Drop out, Skip connection

1. Introduction

In order for the automatic piano player [1] developed in this laboratory to perform with human-like intonation, MIDI (Musical Instrument Digital Interface) standard data with intonation for each note is required. Therefore, inputting score data as it is into the device as-is does not result in human-like performance, which is the objective. Therefore, previous research has developed an inference system using deep learning to represent MIDI-standard data with intonation. However, it was difficult to reproduce the intensity of the sound. To solve this problem, we developed a deep learning model that combines two networks. This paper describes the inference system developed and the inference results.

2. Data used

The training data used was performance data recorded in accordance with the MIDI standard. This data includes parameters of performance information that represent the pianist's performance expression, as well as musical notation information such as notes and dynamics. However, it is difficult to use the MIDI standard data as-is in machine learning, so we extracted the performance information and score information to be used from the performance data and set the parameters so that they can be used in machine learning.

2.1. Performance information

Among the performance data, information that is expressed differently by different pianists is defined as performance information, and four parameters are set. Table 1 shows the performance information.

Table 1 Performance information

Parameters	Detail	Unit
Velo	Sound intensity	-
Gate	Sound length	ms
Step	Interval until next note	ms
Time	Sound time	ms

2.2. *Score information*

We defined score information as information in the performance data that is uniquely determined regardless of the pianist, such as notes and dynamics, and set five parameters.
Table 2 shows the score information.

Table 2 Score information

Parameters	Detail	Unit
Tgate	Sound length on the score	ms
Tstep	Interval until next note on the score	ms
Key	Sound height	-
Bar	Bar number	-
Dyn	Dynamics mark	-

3. Inference System

Fig.1 shows the inference system created for this project. This system was created by combining two networks. Both Network 1 and Network 2 have three intermediate layers. The flow of inference results until they are output is as follows: First, music information is input to Network 1, and then the music information and Velo output by Network 1 are input to Network 2. The final Velo value is obtained from these two networks. Considering the possibility of overlearning, skip connections and dropouts were applied to each intermediate layer.

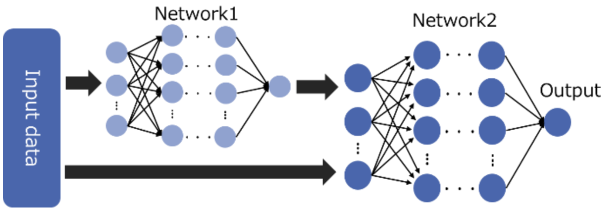

Fig.1 Network layout diagram

4. Experiments

4.1. *Experimental details*

Eight pieces performed by the pianist Vladimir Davidovich Ashkenazy were prepared, one of which was treated as test data, and the remaining seven pieces were used to infer the test data. The actual test data and the inference results are then compared. In this case, Prelude Op.28-7 was used as the test data.

4.2. *Data partitioning for cross-validation*

The hyperparameters of the model were determined by increasing the data set using K-fold cross-validation. By using this method, the evaluation of the model does not depend on the division of the data, and more reliable results can be obtained. Specifically, the eighth song was fixed as the test data, and the other seven songs were divided into five training data and two validation data, for a total of 21 different data sets. The divided data is shown in Table 3.

Table 3 Pattern of data

Pattern number	Train data	Evaluate data	Test data
1	1,2,3,4,5	6,7	8
2	1,2,3,4,6	5,7	8
3	1,2,3,4,7	5,6	8
4	1,2,3,5,6	4,7	8
5	1,2,3,5,7	4,6	8
⋮			
21	3,4,5,6,7	1,2	8

4.3. *Experimental Results*

Fig.2 shows a graph of Velo output by the inference system and Velo in the actual performance data. The graph of Velo output by Network 1 is also shown in Fig.3.

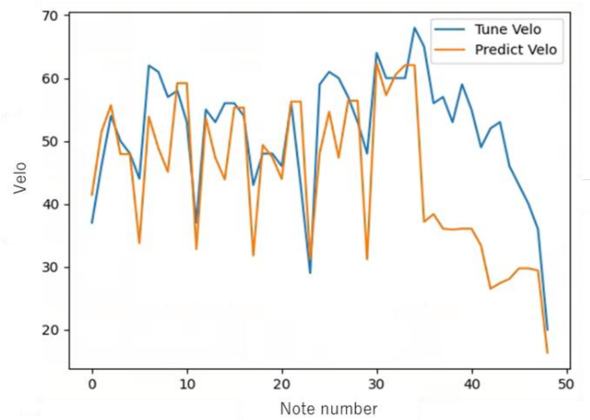

Fig.2 Inference results in the two-stage system

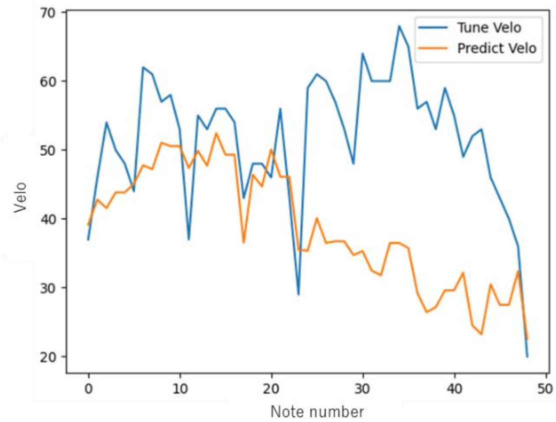

Fig.3 Inference results in the first stage system

5. Consideration

Fig.4 shows the inference results of the previous study. In the previous study, a two-layer inference system was used for the middle layer.

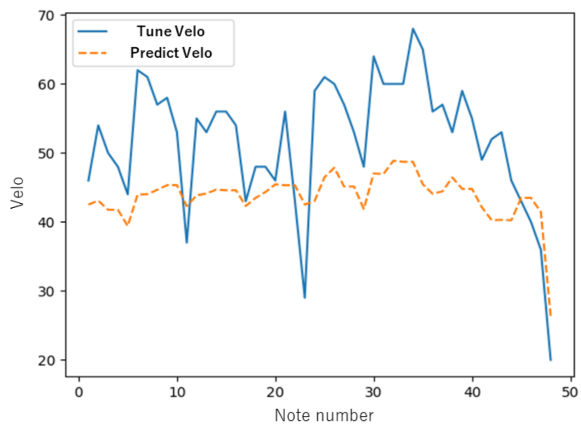

Fig.4 Inference results in previous research

The results of inference using the system combining the two networks in Fig.2 and 4 show that the waveforms of the graphs up to around musical score number 35 are more similar than when inference was performed using the system in the previous study. This is because the features obtained in Network 1 contributed to the output of Network 2, amplifying the change in the waveform, although the change in the graph of Network 1 in Fig.3 is low. However, the waveforms of the graphs after score number 35 are not similar. In particular, the waveforms of Network 1 are less similar to those of the previous study. Network 2 has more correspondence than Network 1, but it does not correspond to the part where the sound suddenly becomes quieter at the end. This may be because the parameters set for the large fluctuation at the end were not appropriate.

6. Conclusion

This study attempted to improve the accuracy of inference of sound intensity by using a model that combines two networks. Experimental results showed that the two network systems produced graphs with waveforms similar to those of the performance data, compared to previous studies. The future prospect is to develop an inference system that can handle large fluctuations. For this purpose, we will verify the improvement of inference accuracy by directly acquiring strength and weakness symbols on the score, which are not included in the MIDI standard data, using image recognition, and inputting them as new parameters into the system.

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. Taiyo Goto

He received bachelor degree in Engineering in 2024 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.

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

Prof. Eiji Hayashi 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).