The research about editing system of performance information for player piano.
-Develop inference methods using machine learning -

Takaaki Ueno
Graduate School of Creative Informatics, Kyushu Institute of Technology
Castail Iizuka 318, 268-1, Kawazu, Iizuka-City, Fukuoka, 820-0067, Japan

Sakmongkon Chumkamon
Department of Mechanical Information Science and Technology, Kyushu Institute of Technology
680-4, Kawazu, Iizuka-City, Fukuoka, 820-8502, Japan

Eiji Hayashi
Department of Mechanical Information Science and Technology, Kyushu Institute of Technology
680-4, Kawazu, Iizuka-City, Fukuoka, 820-8502, Japan

E-mail:
ueno.takaaki568@mail.kyutech.jp

Abstract

In 1996, Professor Eiji Hayashi of this laboratory developed an automatic piano performance system that aims to reproduce performance skills equivalent to those of humans.[1] However, this system does not have a function to infer performance expressions such as the strength and weakness of notes, and is not capable of expressing emotions such as intonation for each phrase, as is done by actual human performers. Therefore, the goal of this research was to reproduce the performances of world-famous pianists under the objective of "pursuit of more human-like performance expression. In the future, when this research matures and the human expression of the piano is clarified mathematically, the data will be used to reproduce human performances, which will fulfill the original purpose of this research. In this paper, we attempted to reproduce "Prelude Op.4 No.10" composed by Fryderyk Franciszek Chopin, using the world-famous pianist "Vladimir Davidovich Ashkenazy" as the research subject.

Keywords: Neural Network, Machine Learning, Long Short-Term Memory, Recurrent Neural Network, Automatic Piano, Computer Music

1. Introduction

In order for an automatic piano to perform in a human-like manner, it needs data on the intonation of each note. However, existing automatic pianos do not have the ability to infer such data. Therefore, until now, data inference has been performed manually. Therefore, we attempted to construct a system that can perform data inference automatically by using machine learning. In this paper, we describe the actual performance information inference system we developed using machine learning.

2. Editing Support System

2.1 Performance Information

The songs shown in Table 1 of CrestMusePEDB, created from MIDI data of a performance by Vladimir Davidovich Ashkenazy (Russia), were used as performance information. Hereafter, the song names will be A to F in Table 1.

Table 1. Information of songs

Sign	musical composition title	Tonality
A	Prelude Op.28 No.4	E minor
B	Prelude Op.28 No.7	A major
C	Prelude Op.28 No.20	C minor
D	Waltz Op.69 No.1	A flat major
E	Waltz Op.64 No.2	C sharp minor
F	Prelude Op.28 No.15	D flat major

2.1 Sheet Music Information

Table 2 shows the variables used in the MIDI (Musical Instrument Digital Interface) standard for musical notation information, where Key represents keyboard position, Velo represents note strength, Gate represents note length, and Time represents playing time.

The information on the strength and weakness of the note, such as forte and piano, is stored in the Dyn variable, and the interval between the next note is stored in the Step variable.

Table 2. Used variable (MIDI)

Parameter	Key	Gate	Velo	Time
Unit	—	ms	—	ms
Reference	21~108	—	1~127	—

2.2 LSTM Model Construction

Long Short-Term Memory (LSTM) was adopted as the inference system for this study. This is an improved model of Recurrent Neural Network (RNN), which is a kind of general circular neural network.

RNN has a continuous relationship between input and output by adding the output of one time to the input of the next time, and can learn data as time-series information by adding the output of each time to the input of each time as memory information. The ability to learn as time-series data is mainly used in the field of natural language processing, such as machine translation of languages. This is because machine translation requires not only word-by-word translation, but also judgment based on the context before and after the translation. The model of LSTM, which is an improved model of RNN, is shown in Figure 1 below.

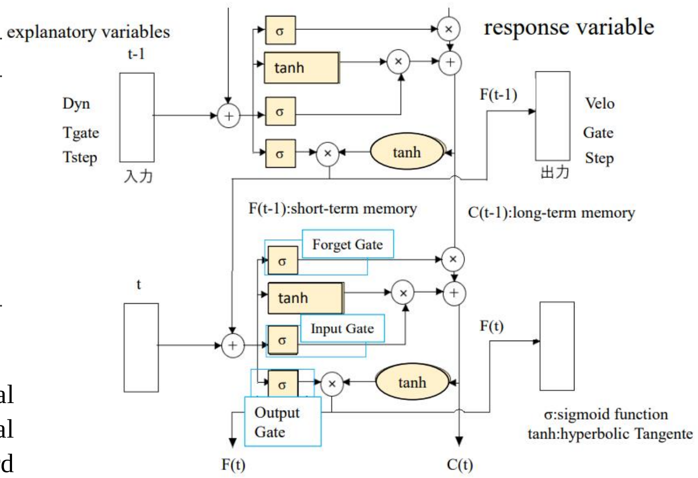

Fig. 1. Abstract of LSTM

LSTM is equipped with a function that divides the given memory information into long-term and short-term memory, and by selecting information at any time and adding important information as long-term memory, it is possible to learn a longer time series than a normal RNN.

3. Inference Experiment

Major headings should be typeset in boldface with the first letter of important words capitalized.

3.1. experimental procedure

As a preliminary step to train the machine learning model, we mapped the created music data to the performance data. 5 songs from A to E were mapped to each other, and then each song was trained for 1000 epochs with 4 songs from A to D as teacher data and E as evaluation data, and inference of F was performed.[2] In learning, Velo, Gate, and Step were placed as objective variables, and Dyn, Tgate, and Tstep, which correspond to the score information of each song, were placed as explanatory variables, judging that they were the most correlated values. After training, the music data of song F, the object of inference, was input to the model, and the performance data obtained as output was evaluated.

3.2. experimental results

Figure 2, Figure 3, and Figure 4 show the results of comparing the inferred values for each of Velo, Gate, and Step with the actual performance (MIDI data).

Fig. 2. Velo comparison of the right hand

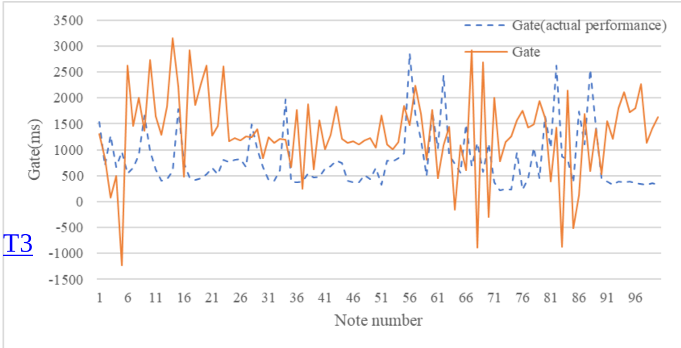

Fig. 3. Gate comparison of the right hand

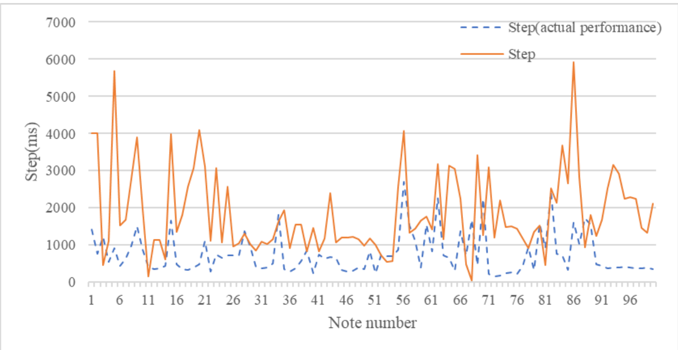

Fig. 4. Step comparison of the right hand

The correlations between the inferred values of Velo, Gate, and Step and the performance (MIDI data) are shown in Table 3, and the loss trends for each piece are shown in Figure 5. The loss function is the mean square error.

Table 3. Correlation with actual performance

	correlation
Velo	0.127
Gate	-0.110
Step	0.245

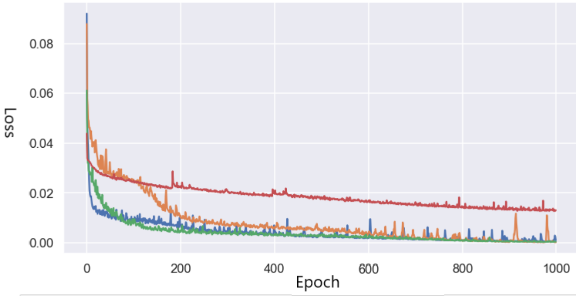

Fig. 5. Loss of 4 songs

4. Consideration

Figure 5 shows that the learning is correct, but Table 3 indicates that the accuracy of inference is low. The fact that few positive correlations were observed suggests that the combination of explanatory and objective variables used as the axis of evaluation for inference was inappropriate.

In addition, in Figure 2, the value of Velo exceeded the upper limit of 127, and in Figure 3, the value of Gate was negative in some cases, indicating that the inference results sometimes included values that were impossible to perform. In addition, abnormal values were output for Step as a performance of about 6000 (ms) in some cases. This may be due to the fact that the model was trained with data that had no output threshold in the model training process, and that data with sharp ups and downs was used as the teacher data.

5. Conclusion

. This year, we constructed a system using machine learning to replace the conventional inference system and conducted inference experiments. This year's experiments made it possible for the system to learn performance information (MIDI data). However, some of the values in the performance inference exceeded the limit of expression, and few positive correlations were observed.

As a future prospect, we aim to construct a system that is capable of sufficient inference. For this purpose, we will review the combination of variables considered in this study, and in addition, we will consider performing inference on other songs and comparing the results.

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.
2. Y. Isomichi, T. Ogawa, Pattern matching by using dynamic programming, J. Inf. Process Soc. Japan 16 (1975), 15–22 (in Japanese).

Authors Introduction

Mr. Takaaki Ueno

He received his B.S. degree in Information Engineering from the Department of Intelligent Systems Engineering, Kyushu Institute of Technology in 2022. He is currently enrolled in the master's program at Kyushu Institute of Technology, where he is conducting research in the Hayashi Laboratory.

Dr. Sakmongkon Chumkamon

Dr. Sakmongkon Chumkamon received Doctor of Engineering degree from Kyushu Institute of Technology in 2017. He was a postdoctoral researcher at Guangdong University of Technology in 2017-2019. Presently he is a postdoctoral researcher in Kyushu Institute of Technology since 2019. His research interests include factory automation robots and social robots.

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).