Relative Representation of Time-Span Tree

Risa Izu[1,*], Yoshinari Takegawa[1] and Keiji Hirata[1]

Future University Hakodate
g2123007@fun.ac.jp

Abstract. We propose a novel representation method of time-span tree of Generative Theory of Tonal Music (GTTM), which is suitable for deep learning using neural networks. We are interested in representing the meaning of music in a tree structure, as in natural language understanding, and employ the time-span tree of GTTM. The strengths of our method are relative tensor representation of parameter values and tree structure of variable shape and size. Our method properly reduces the number of parameter values and the amount of information describing the time-span tree structure for deep learning. That is, the same information can be expressed with fewer symbols. Through small-scale experiments, the relative representation has been shown to be promising.

Keywords: Generative theory of tonal music (GTTM), time-span tree, block view

1 Introduction

Generative theory of tonal music (GTTM) [1] which is a cognitive music theory, represents the hierarchical structure of melodies by expressing the relative importance of each note as a time-span tree. The time-span subtrees exhibit both local and global dependencies, and it is important to consider the both dependencies for a comprehensive analysis of the hierarchical structure of time span trees. Takahashi et al. [2] proposed a method in which a time-span tree is represented by the block view considered as a tensor, and Seq2Seq model with the attention mechanism captures the both local and global dependencies contained in time-span tree. However, since the block view uses absolute values for representing duration and pitch, it leads to difficulty in learning the general rules for the values and the relationships among block.

Therefore, we introduce a block view containing relative values. Specifically, we establish representations for relative vertical and horizontal positions, duration, and pitch, enabling a block view to express the relationships among subtrees. This approach is expected to reduce feature complexity, leading to improved accuracy improvement and reduced of training time.

2 Relative Representation of Time-Span Tree

Our proposed method introduces a relative block view representation, enabling a more detailed and expressive description of the hierarchical structure of melodies. Fig. 1 shows the existing block view converted to a relative representation.

Durations are represented by a combination of nine basic labels, such as quarter note, eighth note, and so on. For example, note id 1 in the 1st layer has a duration of 0.75. Converting this to a relative expression, 0.75 can is represented as the sum of 0.5 and 0.25.

The pitch class is calculated as an interval and direction of melodic change between the pitch and the parent time span that governs the pitch, that is, the block directly superior to the pitch. For example, note id 2 in the 1st layer has pitch class D and is dominated by C♯ in note id 3 in the 2nd layer. D is one interval above C♯, and hence, the interval is 1 and the direction of melodic change is +. In some cases, melodic change may be more than one octave. At present, we assume that melodic change is within one octave (0 to 11) for such cases.

The branching information in the tree structure is represented by the sequence of left- or right-branchings from the maximum time-span position. For the depth of sequence, 0 is assigned to the initial occurrence of time-span (the maximum time-span), and + to the same time-span occurring in the subsequence. Concerning the left/right branching, ϵ is assigned when no branching occurs, and L and R are assigned to the left- and right-branching, respectively. For example, the 4th layer is assigned [0,ϵ] because it has no branches and no upper layers. Furthermore, note id 1 in the 3rd layer is represented as [+, L] because the value 1 means one-level deep from the above 4th layer and left-branching occurs.

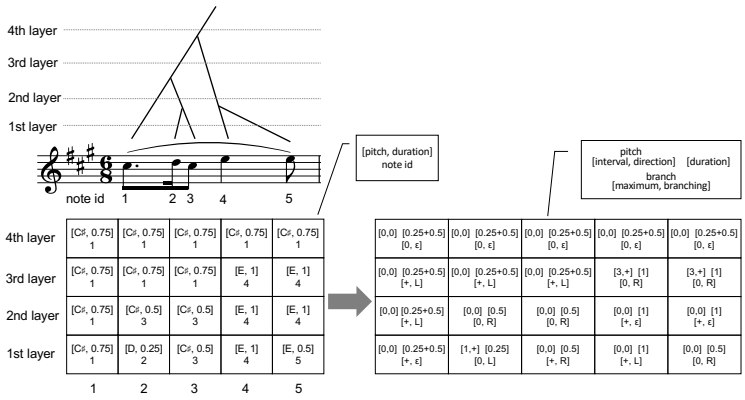

Fig. 1. Conversion of Absolute Representation of Block View to Relative One

When entering data into the model, each note information information is treated as multi-hot. Specifically, we combine a multi-hot vector indicating duration, a one-hot

vector indicating pitch interval, a one-hot vector indicating pitch direction, a one-hot vector indicating branch number, a one-hot vector indicating branch direction, a label indicating padding, a label indicating mask. Table. 1 shows details of each category.

Table 1. Melodic Features in Multi-Hot Vector

Category	Values or Labels	Length
Mask	mask or not	1
Padding	BOS, EOS, padding for sequences, padding for layers	4
Duration	0.125, 0.1667, 0.25, 0.3333, 0.5, 0.6667, 1.0, 2.0, 4.0	9
Pitch interval	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11	12
Pitch direction	0, +, -	3
Sequence of branch	0, +	2
Left/right branching	ϵ, L, R	3

3 Experiment

Through the fill-in-the-blank task for the block view of a time-span tree, let us validate the the proposed relative representation. In this paper, we call the representation method employed in the previous study as the absolute method [2], and the proposed method as the relative method. To evaluate how much the proposed method improves the result of the fill-in-the-blank task over the absolute one, we measure the accuracy of the following three factors: pitch, duration and a pair of pitch and duration. Since a pitch is represented by a pitch interval and the direction of melodic change in the relative method, we have the correct answer if both a pitch interval and the direction are the same.

3.1 Experimental Setup

We use the GTTM database [3], with 176 songs for training data, 44 songs for validation data, and 55 songs for testing data. we crate dataset for the fill-in-the-blank task by masking each subtree. By masking, we obtain 6117 training data, 1498 validation data, and 1797 test data. Each batch contains 64 pieces of data. The embedding dimension by skip-thought is set to 300 and the size of the hidden layer of the Seq2Seq model is set to 200. We use the optimizer Adam with a learning rate of 1.0×10^{-4}.

3.2 Results

Fig. 2 shows the results of the accuracy to masking ratios for the three factors. For pitch accuracy, the relative method exceeded accuracy of 0.60 for almost all masking ratios, and, for all masking ratios, the relative method was superior to the absolute method. For duration accuracy, the absolute method was advantageous for almost all masking ratios. Furthermore, the maximum difference of accuracy rates exceeded 20%. For a pair of pitch and duration, the relative method was equal to or better than the absolute one, and, for low masking ratios, the relative method was superior by about 10%.

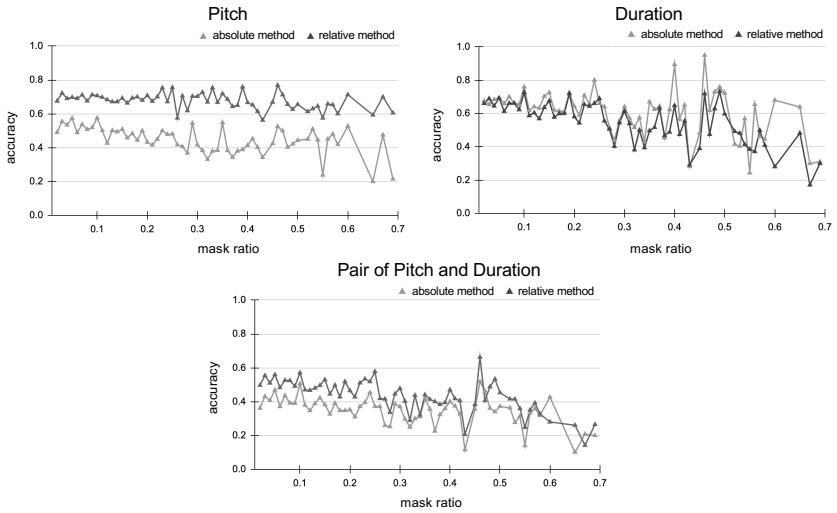

Fig. 2. Accuracy to Masking Ratios for the Three Factors

4 Conclusion

We proposed the relative representation of duration, pitch, and branching information in a block view of time-span tree. The results of the fill-in-the-blank task show that the relative method is advantageous for the factors of pitch and a pair of pitch and duration.

The points to be improved in the future are as follows. To improve duration accuracy, we need to examine and refine the representation method for duration. For example, we consider the relative representation based on metrical information. Furthermore, since the current validation test is conducted on a small dataset, we need to validate on a larger dataset through data augmentation.

References

1. Lerdahl, F., Jackendoff, R.: A Generative Theory of Tonal Music, The MIT Press, Cambridge (1983)
2. Takahashi, R., Izu, R., Takegawa, Y., Hirata, K.: Global Prediction of Time-span Tree by Fill-in-the-blank Task, 16^{th} International Symposium on Computer Music Multidisciplinary Research (CMMR 2023)
3. Hamanaka, M.: GTTM Database, `https://gttm.jp/gttm/ja/database/`