Deep Learning based Detection of GPR6 GTTM Global Feature Rule of Music Scores

Yan-Ru Lai, Alvin Wen-Yu Su

SCREAM Lab., Institute of Computer Science and Information Engineering,
National Cheng Kung University,
No.1, University Road, Tainan City 701, Taiwan (R.O.C.)
{p76081352, alvinsu}@mail.ncku.edu.tw

Abstract. Rules such as phrasing, articulation and intonation are important to musical performance. Local boundaries of a piece of musical score are closely related to these rules. For a computer program performance model, it is desired to detect such local boundaries to manipulate the above rules. The grouping preference rules (GPRs) proposed in the generative theory of tonal music (GTTM) were proven to be effective in this respect. In the past two decades, computer automatic detections of GPRs have been studied. Recently, machine learning techniques were proposed for accurate localization of GPR2 and GPR3. GPR6 is another important feature for local boundary detection. The major difficulties to localize GPR6 lie in the insufficient amount of labeled dataset if deep learning models are used. In this paper, an algorithm is proposed to generate tens of thousands of scores with reliable GPR6 labels. These automatically generated scores are used as the pre-training dataset for the bidirectional long short-term memory (BLSTM) networks. Then, 267 manually labeled data set is used to test the model. The experimental results show that the proposed method is significantly superior to the existing ATTA model.

Keywords. A Generative Theory of Tonal Music, Music Boundary Detection, Grouping Preference Rule, Deep Learning, Bidirectional Long Short-Term Memory Networks

1 Introduction

Since the 20th century, computer science technologies have been applied to many music related researches such as music theories and algorithmic composition [1] and SPEAC [2] proposed by David Cope. In music analysis, the Generative Theory of Tonal Music (GTTM) [3] was proposed and has been used in some applications.

GTTM is a music theory that describes listeners' unconscious understanding of music with four hierarchical structures, such as grouping structures and so on. For grouping structure, the intervals between adjacent groups in the grouping structure is called the local grouping boundary [4]. In order to find the local grouping boundary, GTTM uses the characteristics of musical notes to propose grouping preference rules (GPRs) to help find the local grouping boundary. In deep learning [5] is used to detect local grouping boundaries. Hence, quality labeled datasets are essential. The deep belief networks (DBN) [6] is proposed to detect local boundary, and a model named automatic time-span tree analyzer (ATTA) [7] used to detect GPRs.

In [8], we were able to automatically generate music scores and had trained the model by using these music scores to detect GPR2 and GPR3. Comparing with ATTA, the results of our models outperformed ATTA and reach F-measure over 77% on two thirds of local boundary rules on manual labeled GTTM database [9]. In this paper, the method is proposed to detect GPR6. Because deep learning models are to automatically detect GPR6s, a large amount of training data with labels is required. The GTTM database is not large enough to train our model. Therefore, we propose a procedure that automatically generates labeled scores based on the rules and parameters described in the paper.

To evaluate our method, we generate 10,000 labeled GPR6 music scores to train the model that can detect the GPR6. Since the input data are music scores, we can consider them as temporal sequences of undetermined lengths. Therefore, we use a recurrent neural network (RNN) [10] with a bidirectional long short-term memory (BLSTM) unit [11]. Compared with DBN, the BLSTM networks also provides us with the advantage of being able to handle variable length data. The proposed work is also compared with ATTA. The results show that out model performs better than ATTA on manually labeled GTTM database.

The rest of the paper is organized as follow. In Section 2, brief review of related works is made. Section 3 presents the procedure for algorithmically generating labeled music scores. Section 4 describes the experimental results and some discussions are present. Section 5 gives the conclusion and an overview of future work.

2 Related Works

A Generative Theory of Tonal Music (GTTM)

GTTM [3] is a music theory divides the understanding of music into four hierarchical structures. They are called the grouping structure, the metrical structure, the time-span tree and the prolongation tree, shown in Fig. 1. The grouping structure is known as the most fundamental part of music understanding in GTTM. It divides a piece of music score into many hierarchical groups of notes to simulate how an experienced listener groups musical notes. Moreover, the position between two adjacent groups is called a local boundary. However, there isn't a direct method for the determination of an appropriate grouping structure. In the original GTTM, the grouping well-formedness rules (GWFR) defines all possible structures and the grouping preference rules (GPR) only specify preferred structures. Unfortunately, GWFR and GPR may conflict with each other, though experienced listeners can determine one or more suitable grouping structures. One way to find the grouping structures is to locate the local boundaries in a musical score by using the grouping preference rules in the first place.

There are two types of grouping preference rules. Rules of the first type is called the Local Detail Rules (LDR) and is related to local details such as attack, articulation, dynamics, and registration, which may lead to the perception of group boundaries. In

GTTM， GPR 2a, 2b, 3a, 3b, 3c and 3d are all LDR types. The rule of the second type involves more global considerations such as symmetry and thematic, motivic, harmonic parallelism, rhythmic and so on. For example, GPR 6 represents parallelism. In this work, we will focus on automatic generation of training data used for the training of the machine learning model for GPR 6.

Though GTTM provides these GPRs. for music structure analysis, computer detection of the rules in a piece of musical scores remains a challenge. In ATTA, rule-based algorithms are used to detect the presence of GPRs. In 2016, deep GTTM-I [4] involved DBN (Deep Belief Neural Network) [6] to detect GPRs and local boundaries. In 2020, detection accuracy of GPRs is advanced by using BLSTM (Bidirectional Long Short-Term Memory) neural networks [11].

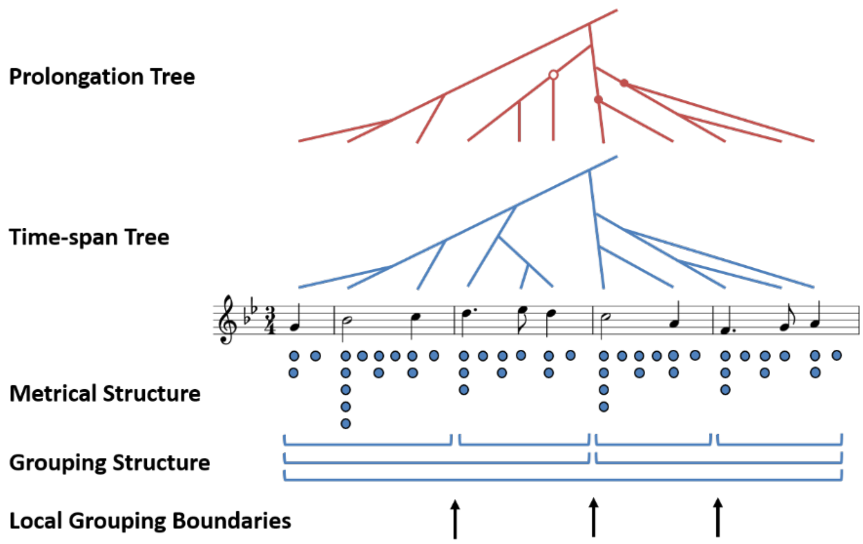

Fig. 1. An Example of the Four Structures and Local Boundaries in GTTM [4].

Annotated Symbolic Music Datasets

Conventionally, symbolic music datasets for various research topics are labeled by human experts. There are datasets in for harmony analysis [12-15], and some are for the training of harmony recognition models [16-18]. In music theory analysis, Schenker41 has performed Schenkerian analyses of 41 excerpts and stored them in a machine-readable format [19]. There are also 300 manually labeled rules and structures in GTTM which is employed in this paper. This GTTM database is also used as a training dataset for DBN to detect the grouping structure and metric structure in [4, 20, 21]. GPR6 is also manually labeled in the GTTM database. In this paper, we use this GTTM GPR6 labels to compare with our own algorithmically generated dataset.

Automatic Time-span Tree Analyzer (ATTA)

ATTA [7] is a rule-based Interactive GTTM Analyzer [22]. ATTA can read a piece of music score in MusicXML [23] format, convert and show them in piano roll. Then, it generates the four structures of GTTM and also marks the local boundary rules with a piano roll tool. There are some limitations when using Interactive GTTM Analyzer

[22]. For example, some music notation information cannot be used when reading MusicXML [23]. We will report this in another paper. Nevertheless, ATTA still provides a feasible reference so that we can verify the performance of our method.

Bidirectional Long Short-Term Memory (BLSTM) neural network

The BLSTM network [11] combines bidirectional recurrent neural network (BRNN) [10] and Long Short-Term Memory (LSTM) cells. BRNN can help the network learn contextual relationships through forward and backward directions. The structure of a LSTM cell is shown in Figure 2, which consists of an input gate i_t, a forget gate f_t and an output gate o_t. LSTM cells improve the lack of long-term memory of RNN, and it provides a flexible and stable way to maintain the long-term dependence of time series. BLSTM networks are used in some music research fields, such as chord generation [24] and harmony recognition [17]. In our previous study [8], the BLSTM network was also used to detect GPR2 and GPR3 with high accuracy.

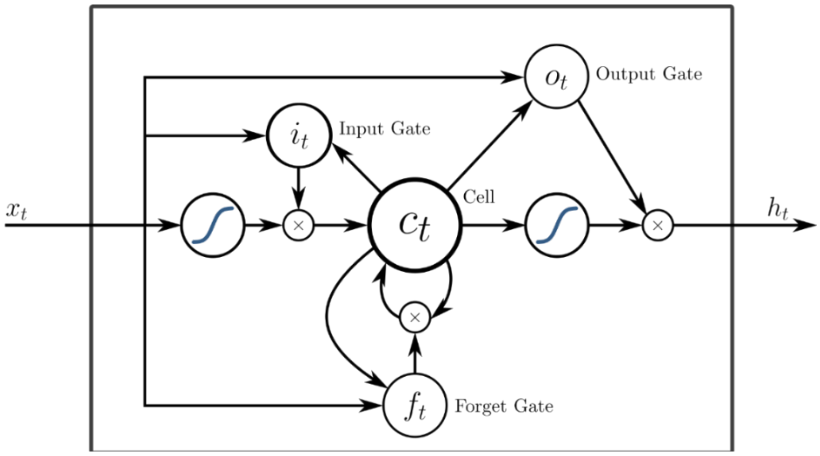

Fig. 2. The structure of a simple LSTM [11].

3 Automatic Generation of Labeled GPR6 Music Score (LGMS)

In order to train a machine learning model to detect the presence of GPR6, we need labeled data to train the model. However, the amount of data in the existing manually labeled data set is insufficient. In this work, a method called LGMS (Labeled GPR6 Music Score) to automatically generate music scores with reliable GPR6 labels is proposed. LGMS consists of several parts, including property initialization, music fragment generation, parallel music fragment generation, score composition of parallel music fragments, and generation of GPR6 labels. All the music scores are generated and stored in MusicXML [23] format by using music21 toolkit [25]. Because the maximum number of measures in the original GTTM database [9] is 16, we also follow this condition in this works though longer music scores can be generated with the proposed method. Same as the GTTM database, all scores are monophonic. Similar to the method mentioned in [8], these automatically generation scores are used to train the BLSTM model.

Property Initialization
In order to create considerable differences in the automatically generated data set, the basic properties of each labeled GPR6 music score must first be initialized as follows:
1) The time signature: 2/4, 3/4 or 4/4.
2) The number of measures of a score: 4, 8, 12 or 16.
3) The key: one of twelve keys without considering its major or minor quality.
4) The shortest length of notes: sixteenth, eighth or quarter note.

Music Fragment Generation

To meet the requirements of GPR6, the properties of each music fragment are as follows:
1) The length of generated segment: A number between 0.5 and 2 multiplied by the time signature's numerator number.
2) 2) A number between 0.4 and 1 multiplied by the maximum number of notes, which can be determined by length of a generated segments as well as the shortest length of notes.

Parallel Music Fragment Generation

In order to generate parallel music fragments, we make the following changes to the previously generated music fragment:
1) Transposed the fragment.
2) Randomly select a few notes in the fragment to make pitch changes.
3) Randomly select long-duration notes from the fragment and split them into several short-duration notes.
4) Randomly select consecutive short-duration notes from the fragment and combine them into a long-duration note.

For example, first randomly generate a two-measure music fragment, as shown in Fig. 3. Next to generate parallel music fragments, we follow the steps below to generate parallel fragments. First, transpose the music fragment. Then in this example, we select the fourth and fifth consecutive eighth notes in the music fragment and combine them into a quarter note as shown in Fig. 4. Merge the two fragments into a new one and label GPR6 which is shown in Fig. 5.

Fig. 3. A generated music fragment.

Fig. 4. A parallel music fragment generated after changes.

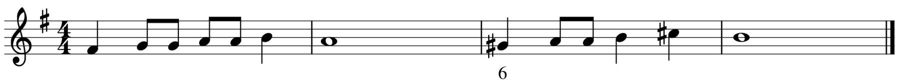

Fig. 5. The new music fragment with a GPR6 denoted below the score.

4 Experimental Results

Feature Representation

In order to detect whether GPR6 exists in the music score, BLSTM networks [11] are used. The input size of the network is 13, which maps to the 13-dimensional features extracted from each note. The first 7 features include the duration, onset, numerator of time signature, beat in measures, duration ratio affected by articulations, pitch and dynamics of each note, and the last 6 features of this note include staccato, staccatissimo, tenuto, accent, strong accent, and under a slur or not.

BLSTM Network Architecture

In our research, the BLSTM network has two stacked layers with a hidden size of 128. The model was trained for 100 epochs and the Adam optimization method of the learning rate is set to 10^{-3}. Batch size is set to 8 scores.

The experiment in detecting GPR6 is divided into two parts. In the first part, 10,000 LGMS are generated. 8,000 of them are used as the training data for the BLSTM network. the rest 2,000 are the validation data. In the second part, the 300 scores of the GTTM dataset having their local boundaries manually labeled by experts are used to test the model trained in the first part.

Validation Result

The results of our verification are evaluated on all intervals and presented in Table I. This result shows that our model has learned how to detect GPR6 from the generated dataset.

TABLE I. VALIDATION RESULT.

Rule	Precision (%)	Recall (%)	F-score (%)
GPR6(Parallel)	95.05%	97.87%	96.43%

Testing Results

In the first 267 scores of the GTTM dataset, all the rules including GPR6 are manually labeled only on the local boundaries. Therefore, the GPR6 detection results of ATTA and our model are compared on the local boundaries for the 267 scores. S1 denotes the case that all 267 scores are used as the testing data. However, it is found that GPR6 labels appear only in 119 scores out of the 267 music scores. S2 denotes the case that these 119 scores are used. In addition, one can manually adjust some parameters to get results when using ATTA. when using ATTA to detect GPR6. Therefore, in order to make the comparison result fair, we will also select the better part of the dataset that ATTA detects GPR6 as our testing set. There are 61 scores in S2 dataset, they can get f-measure greater than 0% by using ATTA. S3 denotes the case that the 61 scores are used. Among the 61 scores, there are 31 scores whose F-score is greater than 50%. S4 denotes the case that the 31 scores are used. All four cases are shown in TABLE II.

TABLE II. PERFORMANCES OF OUR METHOD AND ATTA.

Dataset	Technique	Precision (%)	Recall (%)	F-measure (%)
S1	Our method	51.28%	64.55%	57.15%
	ATTA	21.04%	23.86%	22.35%
S2	Our method	62.55%	64.55%	63.53%
	ATTA	43.20%	23.86%	30.74%
S3	Our method	70.79%	65.72%	68.16%
	ATTA	67.59%	46.79%	55.30%
S4	Our method	75.64%	66.38%	70.70%
	ATTA	87.50%	70.00%	77.78%

Discussion

To summarize the experimental results, we presented four observations. First, in TABLE I, we can find that the BLSTM network achieves 95% F-measure when part of LGMS is used as the validation dataset. Second, when the GTTM dataset is used as the test dataset, the performance of this work reduces which can be seen in TABLE II. There are two reasons make the lower performance, one is that the annotators of the GTTM dataset tend to choose a smaller number of labels, the other is the music parallel fragments we generated are not diverse in our dataset. Third, for the first three testing datasets (S1, S2 and S3), the proposed work performs much better than ATTA. Only in S4, however, ATTA gives slightly better results than the proposed model. Finally, labeling is sometimes subjective. By taking Fig.6 as an example, there is a disagreement between the human annotator and the proposed work. Both consider measure 1-2 and measure 3-4 as respective parallel fragments. The proposed work regards measure 2-3 and measure 4-5 as respective parallel fragments, too. Therefore, there is a GPR6 label under the 4th measure. But the human annotator doesn’t consider that this position has no GPR6. This is because similar patterns can be generated and considered as parallel fragments with the algorithm mentioned in section 3. More researches can be performed in this regard.

Fig. 6. An example in the GTTM dataset. The local boundaries are labeled as "^". The original GTTM labels are labeled in black without parentheses. Our predicted labels are labeled in red.

5 Conclusion

In this work, automatic generation of training data, the LGMS, based on the features of GPR 6 of GTTM is presented. The experiments show that the BLSTM network can be very effective in detecting GPR6 of manually labeled testing database provided by [26] when the above LGMS are used as its training data set. When tested with GTTM database, the proposed work is approximately 34% better than ATTA. In the future, detection of local boundaries by using LGMS will be performed. It is also desired to extend LGMS to other symbolic music information retrieval areas when machine learning models are applied.

References

[1] Muscutt, K., *Composing with algorithms: An interview with David Cope.* Computer Music Journal, 2007. **31**(3): p. 10-22.

[2] Wiggins, G.A., *Computer models of musical creativity: A review of computer models of musical creativity by David Cope.* Literary and Linguistic Computing, 2008. **23**(1): p. 109-116.

[3] Lerdahl, F. and R.S. Jackendoff, *A generative theory of tonal music*. 1996: MIT press.

[4] Hamanaka, M., K. Hirata, and S. Tojo. *deepGTTM-I: Local boundaries analyzer based on deep learning technique*. in *Proceedings of the 12th International Symposium on Computer Music Multidisciplinary Research (CMMR 2016)*. 2016.

[5] Goodfellow, I., Y. Bengio, and A. Courville, *Deep learning*. 2016: MIT press.

[6] Hinton, G.E., *Deep belief networks.* Scholarpedia, 2009. **4**(5): p. 5947.

[7] Hamanaka, M., K. Hirata, and S. Tojo, *Implementing 'A generative theory of tonal music'.* Journal of New Music Research, 2006. **35**(4): p. 249-277.

[8] Xiao, Y.C. and A.W.Y. Su. *Detection of Local Boundaries of Music Scores with BLSTM by using Algorithmically Generated Labeled Training Data of GTTM Rules*. in *International Conference on New Music Concepts (ICNMC)*. 2020.

[9] Hamanaka, M., K. Hirata, and S. Tojo. *Musical structural analysis database based on GTTM.* in *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*. 2014.

[10] Schuster, M. and K.K. Paliwal, *Bidirectional recurrent neural networks.* IEEE Transactions on Signal Processing, 1997. **45**(11): p. 2673-2681.

[11] Hochreiter, S. and J. Schmidhuber, *Long short-term memory.* Neural computation, 1997. **9**(8): p. 1735-1780.

[12] Neuwirth, M., et al., *The annotated beethoven corpus (ABC): A dataset of harmonic analyses of all beethoven string quartets.* Frontiers in Digital Humanities, 2018. **5**: p. 16.

[13] Kaneko, H., D. Kawakami, and S. Sagayama. *Functional harmony annotation database for statistical music analysis*. in *International Society for Music Information Retrieval Conference (ISMIR): Late Breaking session*. 2010.

[14] Devaney, J., et al. *Theme and variation encodings with roman numerals (TAVERN): A new data set for symbolic music analysis*. in *International Society of Music Information Retrieval Conference (ISMIR)*. 2015.

[15] White, C.W. and I. Quinn, *The Yale-classical archives corpus.* Empirical Musicology Review, 2016. **11**(1).

[16] Masada, K. and R.C. Bunescu. *Chord recognition in symbolic music using semi-Markov conditional random fields*. in *International Society of Music Information Retrieval Conference (ISMIR)*. 2017.

[17] Chen, T.P. and L. Su. *Functional harmony recognition of symbolic music data with multi-task recurrent neural networks*. in *International Society of Music Information Retrieval Conference (ISMIR)*. 2018.

[18] Chen, T.P. and L. Su. *Harmony transformer: Incorporating chord segmentation into harmony recognition*. in *International Society of Music Information Retrieval Conference (ISMIR)*. 2019.

[19] Kirlin, P.B. *A data set for computational studies of Schenkerian analysis*. in *International Society of Music Information Retrieval Conference (ISMIR)*. 2014.

[20] Hamanaka, M., K. Hirata, and S. Tojo. *deepGTTM-I&II: Local boundary and metrical structure analyzer based on deep learning technique*. in *International Symposium on Computer Music Multidisciplinary Research (CMMR)*. 2016. Springer.

[21] Hamanaka, M., K. Hirata, and S. Tojo. *deepGTTM-III: Multi-task learning with grouping and metrical structures*. in *International Symposium on Computer Music Multidisciplinary Research (CMMR)*. 2017. Springer.

[22] Hamanaka, M. and S. Tojo. *Interactive GTTM analyzer*. in *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*. 2009.

[23] Good, M., *MusicXML for notation and analysis.* The virtual score: representation, retrieval, restoration, 2001. **12**: p. 113-124.

[24] Lim, H., S. Rhyu, and K. Lee. *Chord generation from symbolic melody using BLSTM networks*. in *International Society for Music Information Retrieval Conference (ISMIR)*. 2017.

[25] Cuthbert, M.S. and C. Ariza. *music21: A toolkit for computer-aided musicology and symbolic music data*. in *International Society of Music Information Retrieval Conference (ISMIR)*. 2010.