MUSIC CLASSIFICATION WITH CONTINUAL LEARNING

Dooyoung Hong
KAIST
dyhong@rit.kaist.ac.kr

Yongjun Kim
KAIST
yongjun@kaist.ac.kr

Inyoung Paik
KAIST
humandream@kaist.ac.kr

ABSTRACT

Artificial intelligence (AI) technology for music shows outstanding performance these days. However, catastrophic forgetting is still having a major negative impact on the development of music technology using AI. Continual learning is a technology that enables neural networks to learn continuously. First, we implement a music classifier according to each genre through a feed-forward model. Then, we apply dynamically expendable networks to the music classifier for overcoming the catastrophic forgetting of neural networks. Finally, we present the lifelong music classifier through a continual learning technique.

1. INTRODUCTION

Music has always been with the development of mankind, and the music industry is getting bigger day by day. In line with that, AI technology for music is also developing. However, catastrophic forgetting, a chronic problem of neural networks that can continuously learn, is still having a major negative affection on the development of music technology using AI.

Catastrophic forgetting, neural networks forget previously learned information as learning continues, is one of the biggest problems with AI toward the next level. Continual learning is a solution that can continuously maintain information in the neural network. It contains many methodologies such as memory replay, weight regularization, and dynamic architecture. They have their own advantages depending on how they are used. However, we deal with dynamic architecture for the music classification in this work.

First, we configure music classifiers along with the state-of-the-art (SOTA) of music auto-tagging algorithms. We classify MagnaTagAtune datasets by each genre and configure the model using convolutional filters and fully connected layers. Then, we combine the classifier models using the dynamic architecture of the continual learning method. We perform the neural selecting and retraining, network expansion, and network split and duplication on the dynamic architecture. Finally, we present the performance of the combined model using continual learning compared with traditional music classification models.

2. RELATED WORK

There have been many attempts to classify music auto tagging. Choi. *et al.* suggested an approach to combining CNN and RNN for music classification [1]. Pons. *et al.* shows that waveform-based models perform better than spectrogram-based models in large data scenarios [3]. Won. *et al.* performed consistent evaluations of different music tag models on various datasets (MagnaTagATune, Million Song Dataset, and MTG-Jamendo) and presented the reference results using common evaluation metrics (ROC-AUC and PR-AUC). [6]

Continual learning contains memory replay, weight regularization, and dynamic architecture methods. Memory replay uses the generator that is possible to create a distribution of data previously used for model training. Shin. *et al.* perfome the modeling using deep generative replay [5]. It contains a generator that can represent the previous training data well using the generative adversarial network (GAN). This method continuously learns about forgettable data through the data generated by the GAN. Weight regularization adds regularization term that slightly updates the important parameters at the previous task when updates new parameter using other data sets. A representative method is Elastic Weight Consolidation (EWC) [2]. Dynamic architecture mitigates the catastrophic forgetting problem through dynamic changes of the network structure for acceptable new tasks. It improves the number of neurons and layers and performs fine-tuning. A representative algorithm is progressive networks [4].

3. IMPLEMENTATION

3.1 Music classifier

Music classifier has important task that can be used in many musical applications such as music search, music recommendation function, and etc. We configure the music classifier using machine learning techniques. Especially, Mel-Frequency Cepstal Coefficients (MFCC) based on speech domain knowledge and these formulas can help to provide machine learning models with features for interpretation of music data.

We form the music classifier using 2D convolutional filter and fully connected layers, as shown Fig. 1. The main function of machine learning is configuring the approximator that can classify the feature of data based on neural structures and actual data. This model has a simple configuration to classify music genres by assembling features in MFCC using CNN and learning FC with those features.

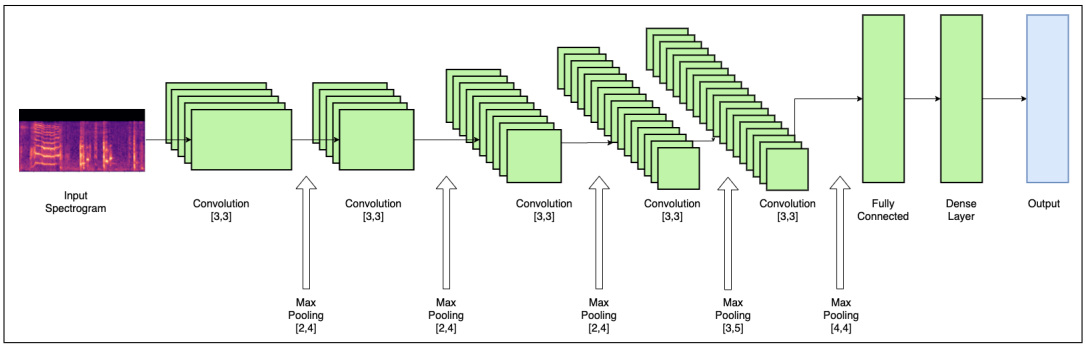

Figure 1. Structure of music classifier model.

We construct subsets of several genres frequently from the MagnaTagATune dataset, and form music classifier using the subset of the data. The configured model has about 75 % of ambiguous performance due to the simple structure. Of course, there are many method to improve the model performance. However, for the easy modification of neurons and layers in the continual learning stage, we use a simple structure as shown now.

3.2 Dynamically expendable networks

Dynamically expandable networks (DEN) proposes the expansion of network's capacity dynamically [7]. DEN aims to goal as following:

- Achieving scalability and efficiency in training.
- Deciding when to expand the network, and how many neurons to add.
- Preventing catastrophic forgetting.

In order to achieve the above goals, DEN performs three sequences of selective retraining, dynamic network expansion, and network split/duplication. Selective retraining retrains by selecting only the parameters relevant to the new task. It only updates parameters related to the discovered task after searching for parameters about a new task. Dynamic network expansion expands the network to reflect new tasks well. It checks whether the model after selective retraining performs well on a new task. The model expands the network dynamically, considering the difficulty of the new task if the model performance is not good. Finally, DEN prevents catastrophic forgetting through network split/duplication. It confirms whether catastrophic forgetting occurs at each neuron and tries retraining after cloning neurons in which catastrophic forgetting has occurred.

3.3 Music classifier with DEN

In progress.

4. AUTHOR CONTRIBUTIONS

"Dooyoung Hong" did procedure the total of storytelling about the Term project. "Yongjun Kim" conducted a survey of the latest technology trends and the distribution of datasets. "Inyoung Paik" implemented the proposed model and performed its validation.

5. REFERENCES

[1] Keunwoo Choi, György Fazekas, Mark Sandler, and Kyunghyun Cho. Convolutional recurrent neural networks for music classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2392–2396. IEEE, 2017.

[2] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114, 12 2016.

[3] Jordi Pons, Oriol Nieto, Matthew Prockup, Erik Schmidt, Andreas Ehmann, and Xavier Serra. End-to-end learning for music audio tagging at scale. *arXiv preprint arXiv:1711.02520*, 2017.

[4] Andrei A. Rusu, Neil C. Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. Progressive neural networks. *CoRR*, abs/1606.04671, 2016.

[5] Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. Continual learning with deep generative replay. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[6] Minz Won, Andres Ferraro, Dmitry Bogdanov, and Xavier Serra. Evaluation of cnn-based automatic music tagging models. *arXiv preprint arXiv:2006.00751*, 2020.

[7] Jaehong Yoon, Eunho Yang, Jeongtae Lee, and Sung Ju Hwang. Lifelong learning with dynamically expandable networks. In *International Conference on Learning Representations*, 2018.