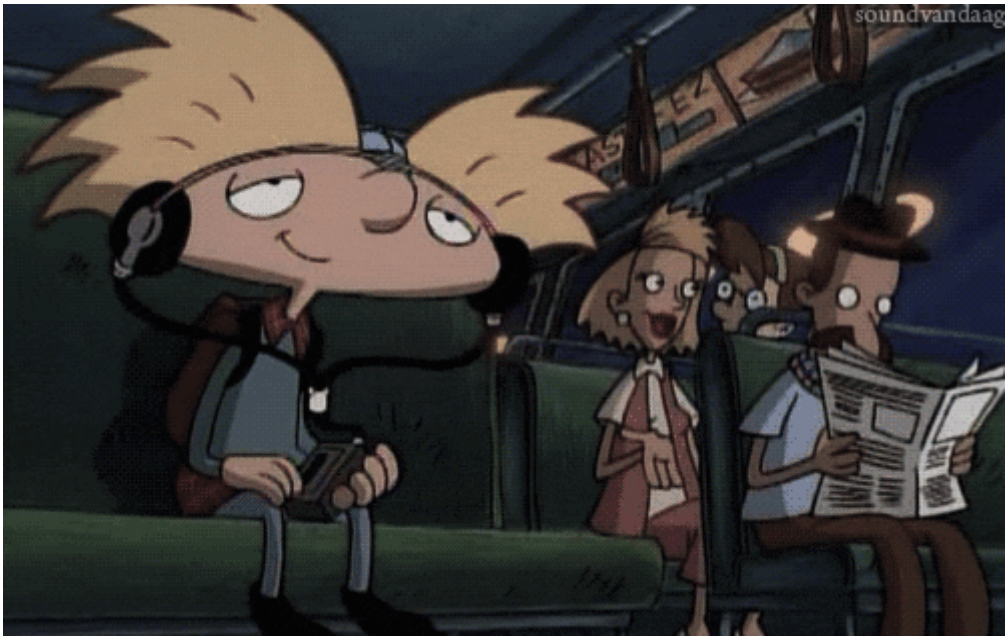
[Paper Summary] A Comparison of Audio Signal Preprocessing Methods for Deep Neural Networks on Music Tagging

Jae Duk Seo
Jun 24, 2018 · 4 min read



GIF from this website

One of my friend was working on audio files with neural networks, and he recommend me to read this paper.

> ***Please note that this post is for my future self to review the materials on this paper without reading it all over again.***

. . .

Abstract

The authors of this paper have performed experiments on music tagging using deep neural networks. They compared different preprocessing methods such as logarithmic magnitude compression, frequency weighting and scaling, and found magnitude compression is the best preprocessing method.

. . .

Introduction

Many of optimizations in machine learning are done via hyper-parameter turnings, however the quality of the input data cannot be ignored. And that applies to audio

'log(X+alpha) where alpha can be arbitrary constants such as very small number (e.g. 10e-7) or 1'

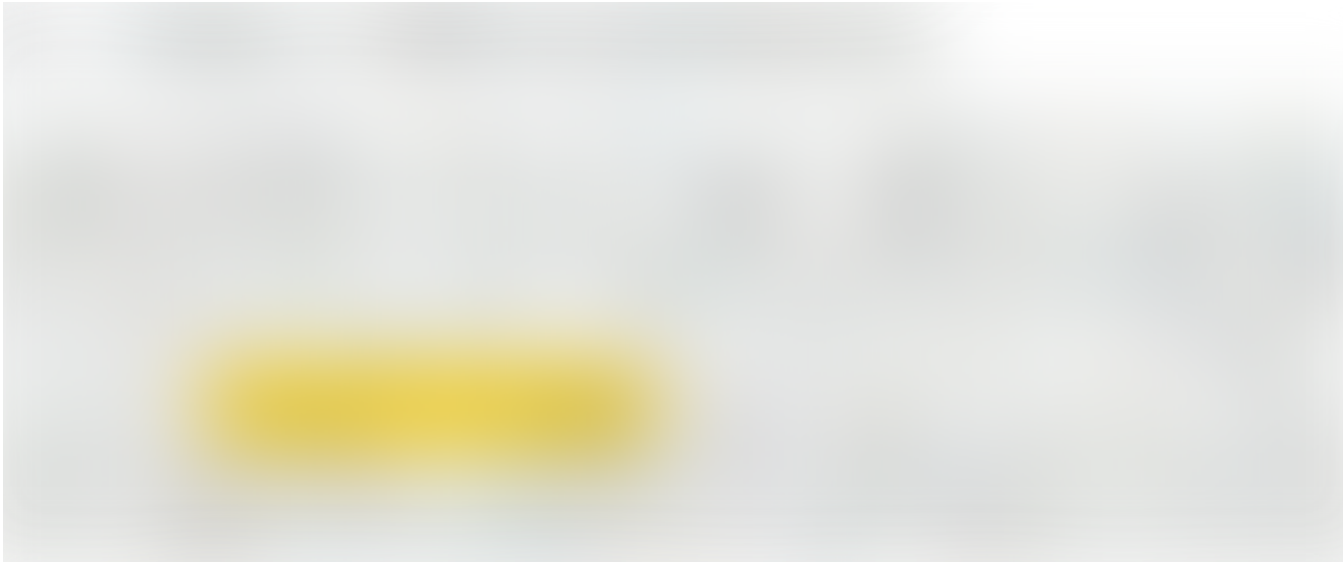
to other preprocessing techniques.

. . .

Experiment and Discussions

In this section, the authors describe the network architecture as well as the input data structure. In short, they used a convolutional neural network with ELU activation, and the input data had a dimension of (1,96,1360). Also they got the music data from Million Song Dataset and to transform the audio into 96 * 1360 dimension they used a discrete Fourier transform. (using the python library Kapre and LibROSA.)

2.1. Variance by different initialization



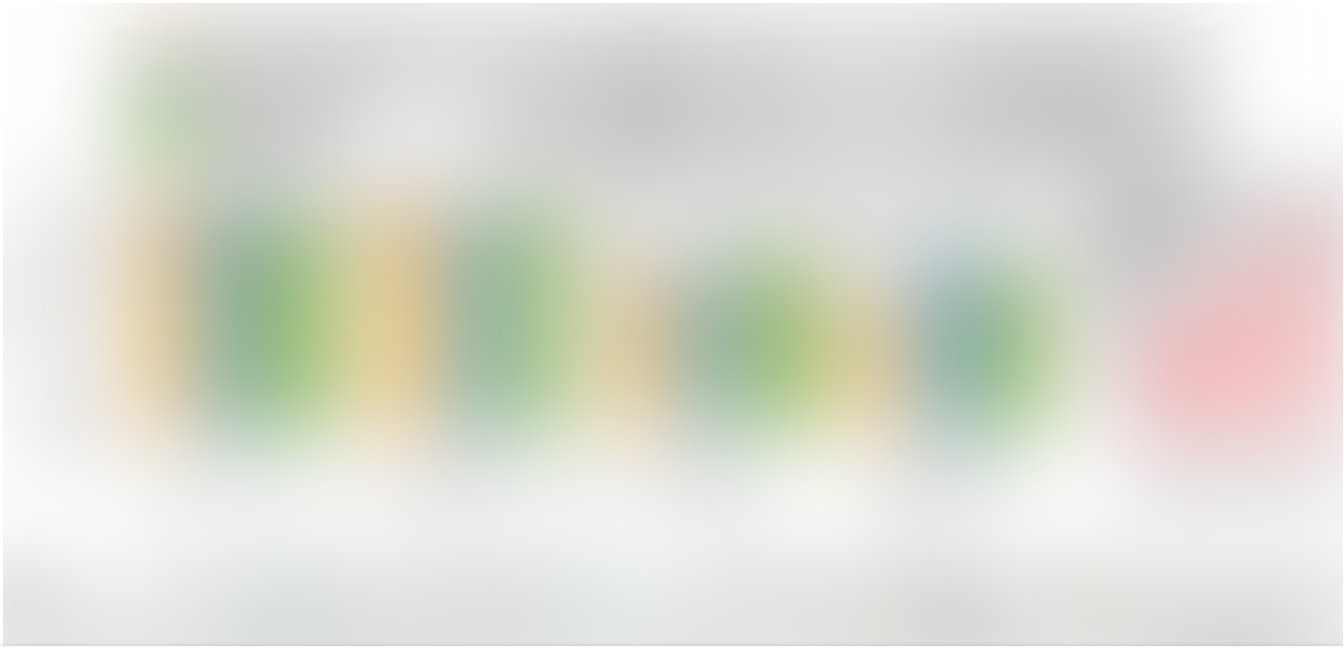
Here the authors have described the fact that they didn't choose to use k-fold cross-validation rather, they repeated the experiments 15 times and compared the AUC scores at each experiments.

• • •

STFT and melspectrogram are most popular method of representing input data in audio classification. (It is generally believed that melspectograms are a better choice with smaller dataset.) However, when the authors performed varieties of experiences with these two preprocessing methods, they found that it wasn't the case for them.

• • •

In this section the authors have experimented with two different input representations log-melspectrogram and melspectrogram, with three frequency weighting schemes per-frequency, A-weighting and bypass, as well as two scaling methods X10 (on) and X1 (off).

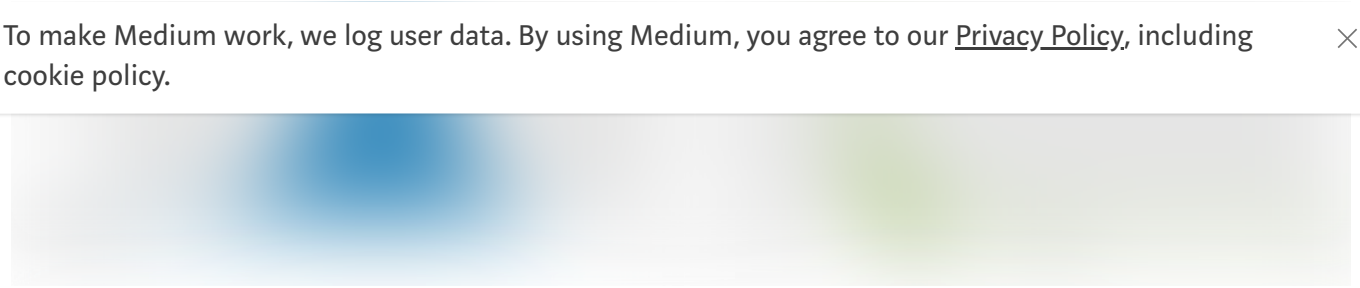


And as seen above, when preprocessing the audio with log() function (either with or without scaling factor of 10, we can observe an increase in the AUC score.

. . .

2.4. Log-compression of magnitudes

One of the reason why applying the log() function to the audio file is a great idea is because it changes the distribution of the data into Gaussian distribution. As seen below, we can observe a smooth bell curve when applying a log() function to the melSpectrogram.



. . .

Conclusion

The authors have found that logarithmic scaling of the magnitude is the best preprocessing method for music classification task.

. . .

Final Words

It was quite an interesting read.

Meanwhile follow me on my twitter here, and visit my website, or my Youtube channel for more content. I also implemented Wide Residual Networks, please click here to view the blog post.

. . .

Reference

1. Choi, K., Fazekas, G., Cho, K., & Sandler, M. (2017). A Comparison of Audio Signal Preprocessing Methods for Deep Neural Networks on Music Tagging. Arxiv.org. Retrieved 23 June 2018, from https://arxiv.org/abs/1709.01922
2. What are A, C & Z Frequency Weightings? — NoiseNews. (2011). NoiseNews. Retrieved 24 June 2018, from https://www.cirrusresearch.co.uk/blog/2011/08/what-are-a-c-z-frequency-weightings/
3. Song Metadata and Why Its Hidden in Your Digital Music Files. (2018). Lifewire. Retrieved 24 June 2018, from https://www.lifewire.com/what-is-music-tagging-2438569
4. Million Song Dataset | scaling MIR research. (2018). Labrosa.ee.columbia.edu. Retrieved 24 June 2018, from https://labrosa.ee.columbia.edu/millionsong/
5. LibROSA — librosa 0.6.1 documentation. (2018). Librosa.github.io. Retrieved 24 June 2018, from https://librosa.github.io/librosa/
6. keunwoochoi/kapre. (2018). GitHub. Retrieved 24 June 2018, from https://github.com/keunwoochoi/kapre
7. Short-time Fourier transform. (2018). En.wikipedia.org. Retrieved 24 June 2018, from https://en.wikipedia.org/wiki/Short-time_Fourier_transform
8. MelSpectrogram. (2018). Fon.hum.uva.nl. Retrieved 24 June 2018, from http://www.fon.hum.uva.nl/praat/manual/MelSpectrogram.html