MIR PROJECT'19: INTRA-FRAME INSTRUMENT CLASSIFICATION

Enric Gusó Muñoz
Univeirsitat Pompeu Fabra
enric.guso01@estudiant.upf.edu

Sergio Gor Bravo
Universitat Pompeu Fabra
sergio.gor01@estudiant.upf.edu

ABSTRACT

In the following report we summarize the work we did in this second LAB.

1. INTRODUCTION

2. METHODOLOGY

3. RESULTS

3.1 SVM

List of features (from most important to less important):

1. MFCCs 2nd to 7th
2. Pitch salience
3. Spectral flatness [in dB]
4. Spectral kurtosis
5. Spectral skewness

k_best	acc%
1	54.38
3	61.25
5	69.01
10	x

Table 1. Validation accuracy using subsets of features in the SVM.

3.2 DNN

#	**model**	**1+log**	**norm**	**drp**	**lr**	**wlen**	**acc%**
1	mlp	no	std	no	E-4	120	79.91
2	mlp	**yes**	**std**	no	**E-4**	120	**80.34**
3	mlp	yes	none	no	E-4	120	79.79
4	mlp	yes	m_std	no	E-4	120	79.80
5	mlp	yes	std	no	E-2	120	29.33
6	mlp	yes	std	no	E-3	120	79.65
7	mlp	yes	std	no	E-4	120	80.38
8	mlp	yes	std	no	E-5	120	80.18
9	mlp	yes	std	0.5	E-4	120	81.97
10	mlp	yes	std	0.5	E-4	30	78.62
11	mlp	yes	std	0.5	E-4	240*	80.98
12	mlp	yes	std	0.5	E-4	513*	81.92
13	mlp	yes	std	0.5	E-4	1024*	81.99
14	**unet**	yes	std	**no**	E-4	120	**82.87**

Table 2. Hyperparameter search results. Best Validation Accuracy among 10 epochs, using 80% of the data for training. *mstd* corresponds to -mean/std standarization, *drp* to dropout, *wlen* to the number of STFT frames assigned to the input layer and *lr* to learning rate. **Those frames are too large to be considered intra-frame classification*.

#	**model**	**augm**	**bnor**	**drp**	**bsize**	**wlen**	**acc%**
14	unet	no	no	no	30	120	82.87
15	unet	yes	no	0.5	30	120	81.41
16	unet	yes	**yes**	0.5	30	120	81.98
17	unet	yes	no	no	30	120	81.85
18	unet	yes	yes	**x2***	30	120	**82.89**
19	unet	yes	yes	x2*	60	120	82.29
20	unet	yes	yes	x2*	15	120	81.49
21	unet	yes	yes	x2*	120	120	82.76

Table 3. Hyperparameter search results. Best Accuracy among 25 epochs, using 80% of the data for training. *mstd* corresponds to -mean/std standarization, *drp* to dropout on the first three layers, *wlen* to the number of STFT frames assigned to the input layer and *lr* to learning rate.

4. DISCUSSION

5. CONCLUSIONS

6. REFERENCES

[1] Bogdanov, Dmitry and Wack, Nicolas and Gómez Gutiérrez, Emilia and Gulati, Sankalp and Herrera Boyer, Perfecto and Mayor, Oscar and Roma Trepat, Gerard and Salamon, Justin and Zapata González, José Ricardo and Serra, Xavier. "Essentia: An audio analysis library for music information retrieval," *14th Conference of the International Society for Music Information Retrieval (ISMIR)*, pp.493-8, 2013.