Song project

EQ2341 Pattern Recognition and Machine Learning

Oriol Closa (oriolcm@kth.se)
Clara Escorihuela Altaba (claraea@kth.se)

May 24th, 2021

Dataset
Samples for training and testing

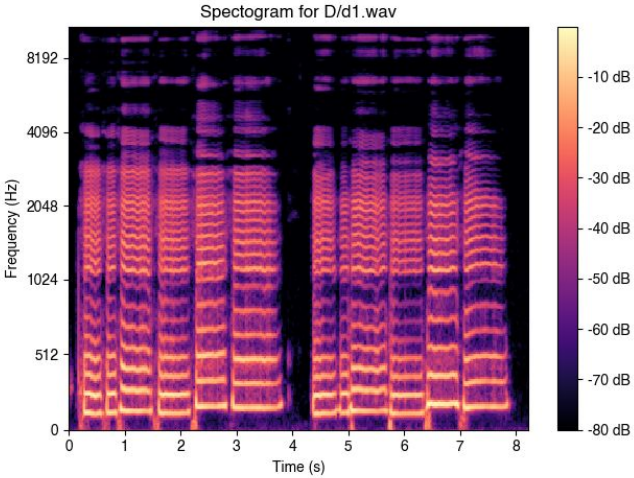
Different large datasets researched.
- Choral singing (ICMPC/ESCOM, 2018).
- Standford's DAMP (ICASSP, 2018).

Finally, we used of our own data and voices to train.

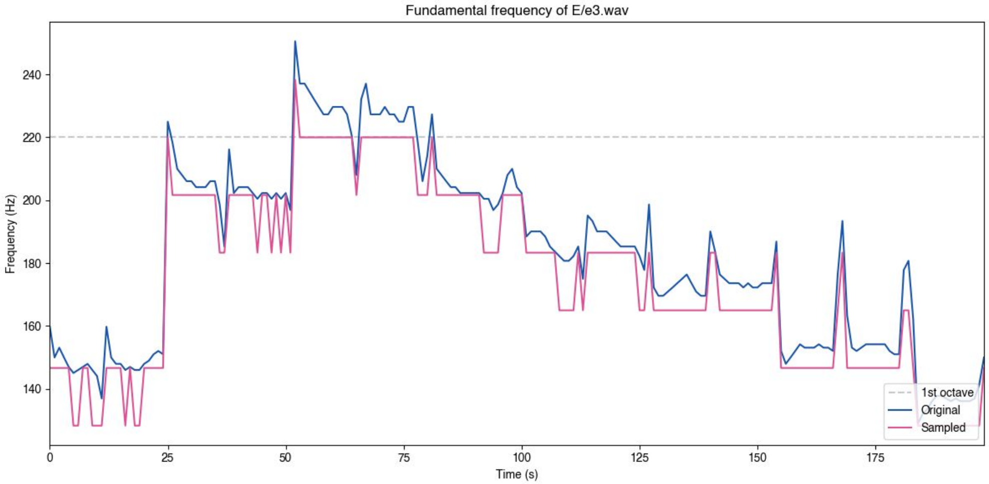
		Training		Test	
		Oriol	Clara	Oriol	Clara
Melody A	Cherry lady	5	0	2	0
Melody B	Happy birthday	6	6	1	1
Melody C	Quan les oques van al camp (traditional Catalan song)	6	6	1	1

Dataset
Spectogram and frequency analysis

1. Apply Yin algorithm to detect note frequency.
2. Infer semitone and octave from note frequency.



Spectogram for Melody B



Frequencies for Melody C

Feature extractor
Theoretical interpretation

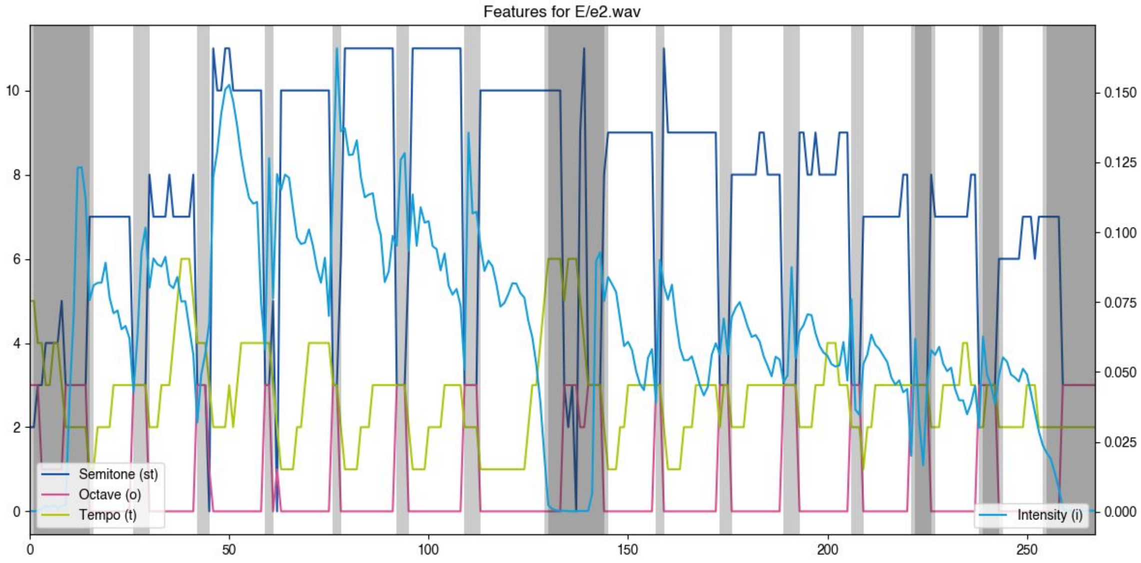
The most relevant features in melodies are timbre, rhythm and dynamics[1]. According to that, our feature extractor contains 8 different parameters.

- Semitone (`st`).
- Octave (`o`).
- Silence (`s`).
- Filtered Silence (`sf`).
- Intensity (`i`).
- Tempo (`t`).
- Semiton Difference (`dst`).
- Octave Difference (`do`).

$$F = \begin{pmatrix} st_1 & st_2 & \cdots & st_n \\ o_1 & o_2 & \cdots & o_n \\ s_1 & s_2 & \cdots & s_n \\ sf_1 & sf_2 & \cdots & sf_n \\ i_1 & i_2 & \cdots & i_n \\ t_1 & t_2 & \cdots & t_n \\ dst_1 & dst_2 & \cdots & dst_n \\ do_1 & do_2 & \cdots & do_n \end{pmatrix}$$

[1] NAWAZ, Rab; NISAR, Humaira; YAP, Vooi; TANG, Py. Acoustic Feature Extraction from Music Songs to Predict Emotions Using Neural Networks. In: 2018, pp. 166–170. Available from DOI: 10.1109/ICBAPS.2018.8527414

Feature extractor
Graphical interpretation



$$F = \begin{pmatrix} st_1 & st_2 & \cdots & st_n \\ o_1 & o_2 & \cdots & o_n \\ s_1 & s_2 & \cdots & s_n \\ sf_1 & sf_2 & \cdots & sf_n \\ i_1 & i_2 & \cdots & i_n \\ t_1 & t_2 & \cdots & t_n \\ dst_1 & dst_2 & \cdots & dst_n \\ do_1 & do_2 & \cdots & do_n \end{pmatrix}$$

HMM implementation

Design, training and prediction

Design

2 different approaches.

a) Discret observation probability matrix ($\lambda = \{\{q, A\}, B_{discret}\}$).

b) Continuous observation probability matrix ($\lambda = \{\{q, A\}, B_{Contininious(GMM)}\}$).

Training

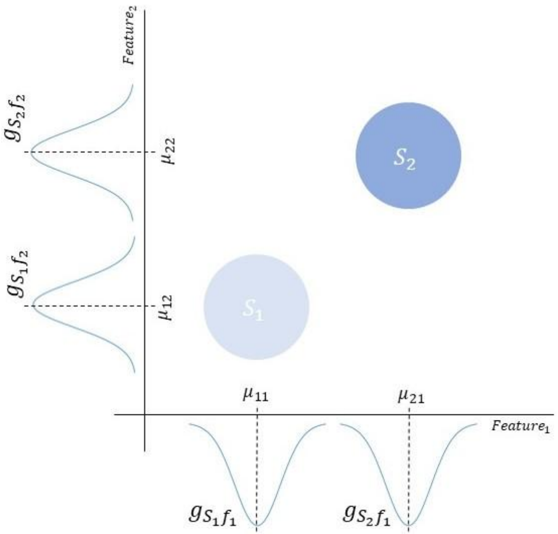
1. Baum-Welch algorithm per song.

Prediction

1. Forward algorithm.
    a. Calculate the `logprob(obs)` per class given the obs sequence.
    b. Select the maximum probability.

Continuous observation probability matrix approach
Gaussian Mixture model

General idea



Characteristics

$$\lambda = \{\{q, A\}, B_{Continious(GMM)}\}$$

$$q_j = [P_1 = j] \approx \frac{1}{N} + \mathcal{N}(\mu, \sigma^2)$$

$$a_{ij} = [P_t = i | P_{t+1} = j] \approx \frac{1}{N} + \mathcal{N}(\mu, \sigma^2)$$

$$b_{i(x_t)} = f_{Xt|S_t}(x_t|i) = \sum_{m=1}^{M} w_{im} g(x_t, \mu_{im}, C_{im})$$

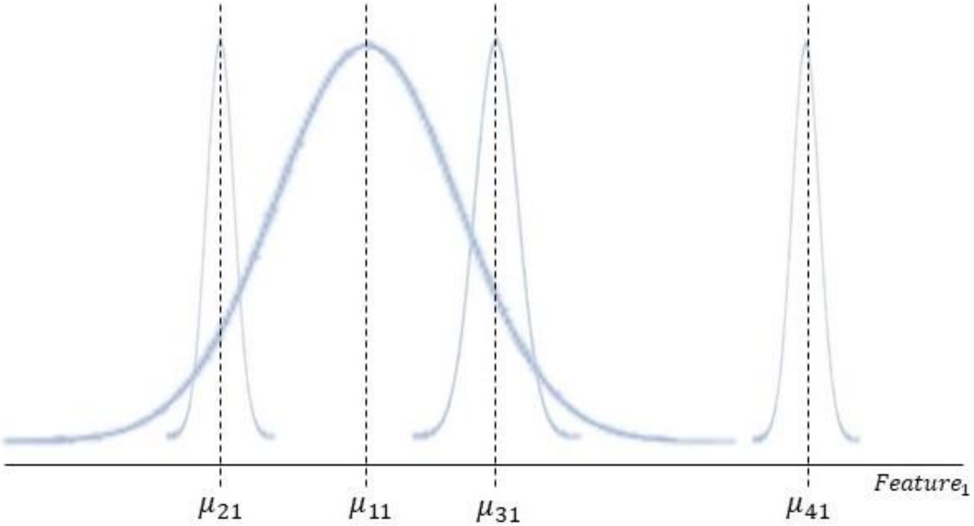
Making sure the transition matrix is row-stochastic.

- N = 5.
- M = 2.
- f_1 = Semitones difference.
- f_2 = Octaves.

Continuous observation probability matrix approach

[8]

Problem with the Gaussian Mixture model



$$\mu_1 = 7; {\sigma_1}^2 = 23$$
$$\mu_2 = 6.8; {\sigma_2}^2 = 0.3$$
$$\mu_3 = 7.5; {\sigma_3}^2 = 0.2$$
$$\mu_4 = 11; {\sigma_4}^2 = 0.1$$

Discret observation probability matrix approach
Theoretical model

Characteristics

$$\lambda = \{\{q, A\}, B_{discret}\}$$

$$q_j = [P_1 = j] \approx \frac{1}{N} + \mathcal{N}(\mu, \sigma^2)$$

$$a_{ij} = [P_t = i | P_{t+1} = j] \approx \frac{1}{N} + \mathcal{N}(\mu, \sigma^2)$$

$$b_{jm} = P[Z_t = m | S_t = j] \approx \frac{1}{M} + \mathcal{N}(\mu, \sigma^2)$$

Making sure the transition and observation matrices are row-stochastic.

- N: 6.
- M: 1 feature for 13 discrete values.
- f_1: Semitones restricted to one octave [-12, 12].

Experimental results
Example of trained matrices

Q matrix

	0	1	2	3	4	5
0	0.00000	0.99675	0.00000	0.00000	0.00000	0.00000

A matrix

	0	1	2	3	4	5
0	0.00000	0.00000	0.79644	0.00000	0.20317	0.00039
1	0.11479	0.27034	0.00000	0.00000	0.61487	0.00000
2	0.00000	0.00005	0.00000	0.00016	0.00000	0.99979
3	0.00002	0.80910	0.02846	0.16240	0.00000	0.00002
4	0.53532	0.21207	0.00000	0.01546	0.00000	0.23715
5	0.05979	0.07119	0.00000	0.19207	0.00000	0.67695

B matrix

9	10	11	12	13	14	15
0.00000	0.03862	0.10082	0.00000	0.06458	0.35102	0.18253
0.07214	0.04362	0.00000	0.36260	0.39507	0.00000	0.00000
0.04765	0.06829	0.23371	0.07918	0.26519	0.00000	0.00000
0.00000	0.00000	0.12102	0.65711	0.09280	0.00000	0.00000
0.06854	0.11026	0.52448	0.08619	0.00001	0.00000	0.00000
0.01058	0.00131	0.00000	0.94460	0.03233	0.00000	0.00000

Results
Description of the tests

We tried multiple different configurations embracing different permutations with the following values.

- N = 2, 3, 4, 6, 8 and 10.
- M = 13, 49, 73 and 169.
- Features: Semitones and semitones difference.
- Data: All columns and only those where there is a change.
- Repetitions: 1, 2, 5 and 100.

Results

Log probabilities of the test songs

	Melody A		Melody B		Melody C	
	Test 1 ✅	Test 2 ❌	Test 1 ✅	Test 2 ✅	Test 1 ✅	Test 2 ✅
Melody A	**-144,50**	-175,05	-315,05	-290,09	-250,94	-212,03
Melody B	-155,01	-152,32	**-286,38**	**-203,31**	-160,26	-181,56
Melody C	-160,45	**-151,23**	-288,26	-207,40	**-155,46**	**-180,32**

`N = 4`, `M = 13`

Discussion
Conclusions and improvements

Conclusions

1. HMMs for song due to high accuracy.
2. Increase dataset of melody A to train more HMM A to get a higher accuracy.

Future improvements

1. Better feature extractor.
2. Apply observation continuous matrix with GMM.

Song project

EQ2341 Pattern Recognition and Machine Learning

Oriol Closa (oriolcm@kth.se)
Clara Escorihuela Altaba (claraea@kth.se)

May 24th, 2021

TITLE

SUBTITLE