Music Information Retrieval-- Homework 1 Report

Key-finding algorithms: global key and local key detection

Student ID: 103062372
Name: 蕭子聲

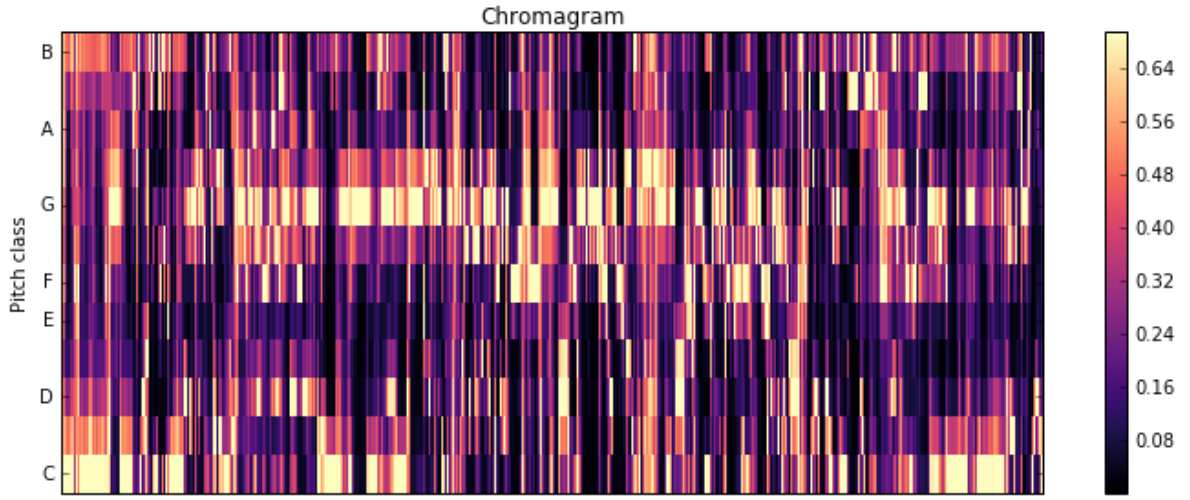
1、 Q1 Discussion

Genre	Accuracy
Total Accuracy	19.39%
Pop	35.10%
Blues	4.08%
Metal	19.35%
Hiphop	9.87%
Rock	27.55%

As the results of the above, we can see that the accuracy of the genres of 'blues' and 'hiphop' are obviously lower than other 3 genres, and the genre 'blues' have the lowest accuracy. I think that's because the authors of these two genres don't like to use the tonic pitch in their song. This leads to the value of the tonic pitch lower than others.

Another reason is that there is no any insurance that the chroma, the most frequently appeared, is the tonic pitch of the song. Especially we only have 30 seconds, which is a vary small segment of the song.

From the example (please refer to jupyter notebook: In [16]), we can definitely see that the key G is the most frequently appeared key, so we will guess G is the right tonic of this song. However, the answer is E minor, which is highly irrelevant to the key G.



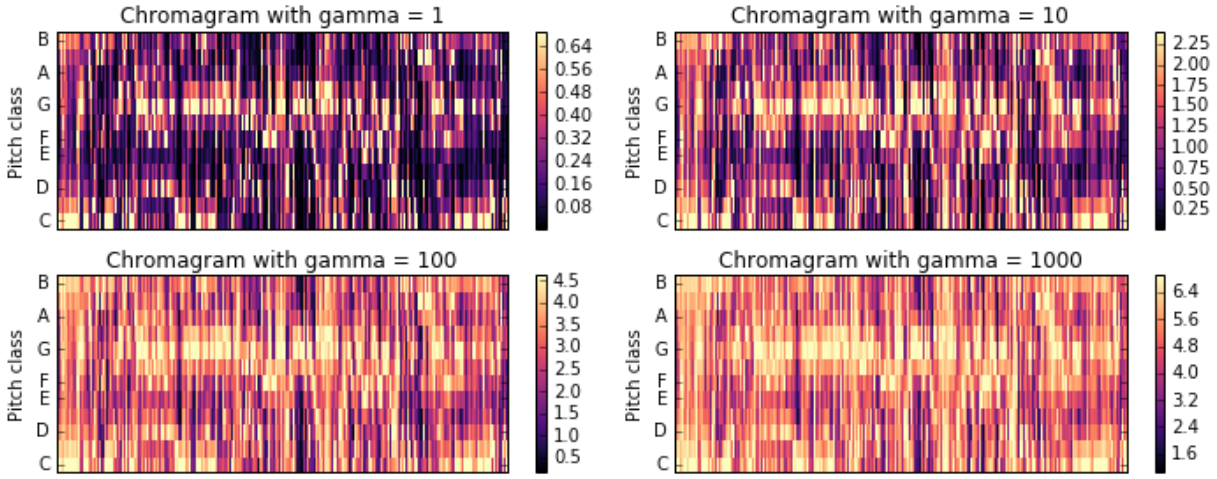
▲ The chromagram of song ‘blues.00001’

2、 Q2 Discussion

Gamma	Accuracy
1	23.70%
10	21.98%
100	19.39%
1000	19.39%

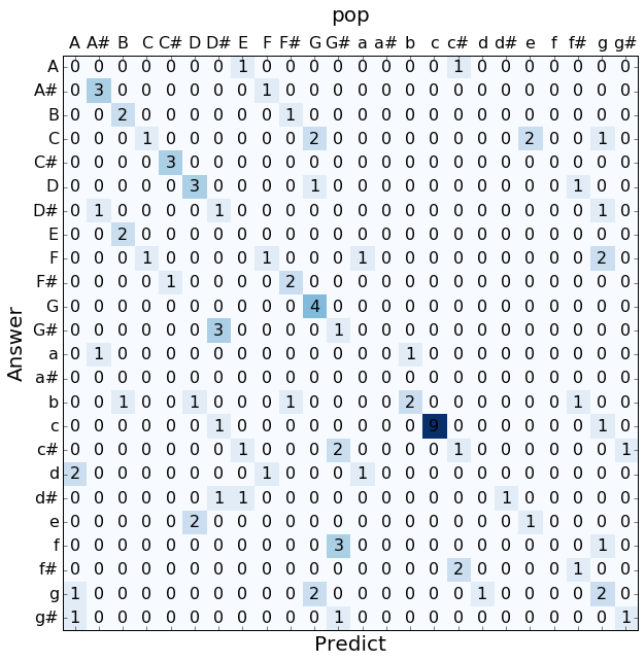
The highest accuracy (23.70%) occurred when gamma = 1.

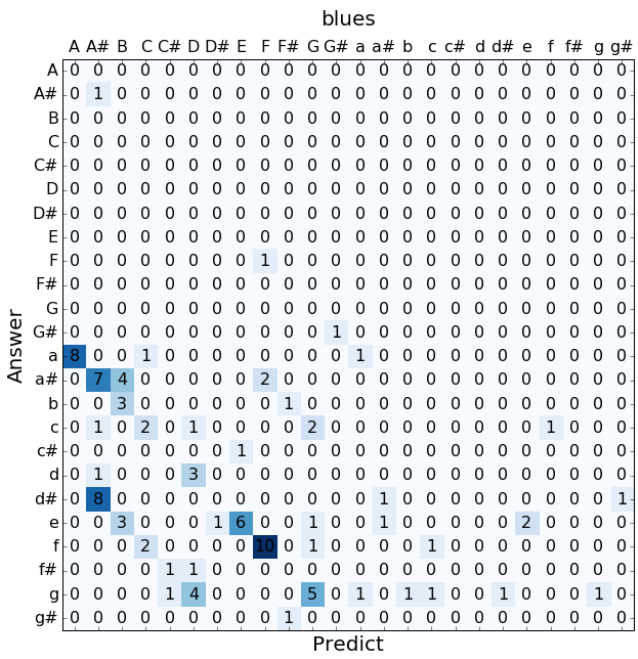
This phenomenon is produced by the features of the logarithm. Since the logarithmic scale would reduce the interval for higher values. Hence, is we use bigger gamma value, $1 + \gamma|x|$ term will scale up the original feature values, and $\log(1 + \gamma|x|)$ term will reduce the differences between each feature values. So if we increase the gamma value, it will become more difficult to detect the features.



3、Q3 Discussion

Gamma	Accuracy
1	33.55%
10	31.72%
100	29.20%
1000	29.03%



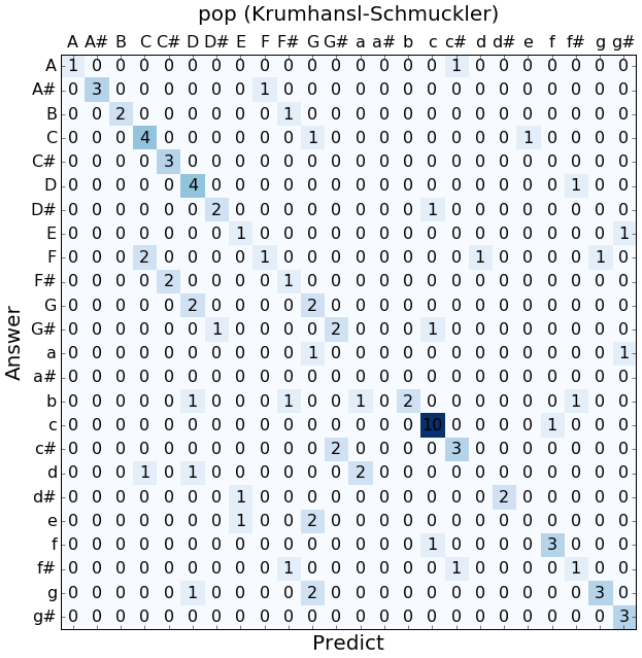


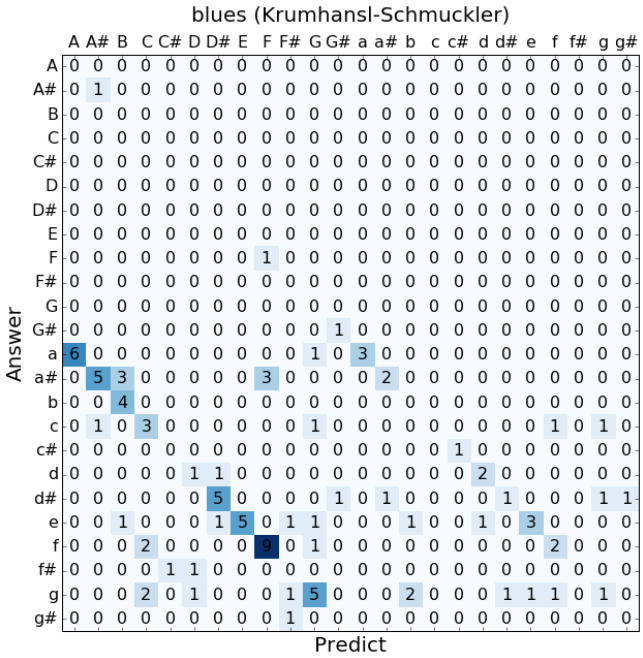
The above two figures show the detection results of the best and worst genres, ‘pop’ and ‘blues’, respectively. The y-axis is the answer key of the song, and the x-axis is what our model predicted.

We can discover that for ‘pop’ genres, there are two lines clearly existing, both are the relation of ‘Same’ and ‘Perfect fifth’ to the correct key. For ‘blues’ genre, the clearest lien is the relation of ‘Parallel major/minor’ to the correct answer. So, if we treat those relations as a part of the correct answer, then we can get higher accuracy.

4、 Q4 Discussion

Gamma	Accuracy (Same)	Accuracy(+5th, rel, pal M/m)
1	33.83%	45.02%
10	32.11%	42.08%
100	31.46%	41.27%
1000	30.81%	40.58%





In this task, we use Krumhansl-Schmuckler's profile, which slightly improved our performance. We can see that the biggest difference between Task2 and Task1 is that the result of 'blues' genre in the Task2 hit more answers than Task1. That's because Krumhansl-Schmuckler use 'Weighted' template to detect the key, which will enhance the feature differences between major and minor key. So the 'Parallel major/minor' relations decreased and the correct answer increased.

The point of the major/minor key feature enhancement of the Krumhansl-Schmuckler's profile is that Krumhansl intuitively increased the weight of 'median key', which is an important feature to distinguish whether a song is in major key or minor key.

5、 Q5 Discussion

- Question 1:

  What is the limitation of these two methods in key detection?

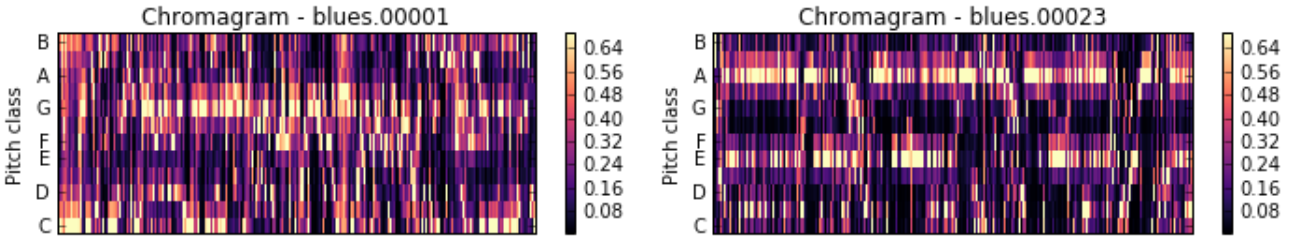
- Answer 1:

  There are two limitations of these two methods:

When we use these methods, we assume that the tonic pitch will appear the most frequently. This will result in some serious problems if the song use other pitches more than tonic pitch.

The following example 1 is definitely violated our assumption. The maximal value in the chroma vector is key G, in contrasts, the minimal value is key E. However, the key E IS the tonic pitch of the song. In this case, we misjudgement the tonic pitch in the first stage, not to mention to find out the major/minor key of this song.

Another example 2 shows that 'blues' songs often use some tricks, for example, periodically switching between major and minor key in the middle of the song with the same tonic. This will seriously frustrate our detection performance. Even if we detected that the major mediant is more than minor median, the song might still be in minor key. Actually, these cases are very difficult to distinguish even for humans.



Another limitation is that summing up chromagram will destroy the over time changes of the features, so this is not suitable to use to detect the local key.

- Question 2:

  Is there any drawback of using the GTZAN dataset for key detection?

- Answer 2:

  Yes, there is. In my opinion, 30 seconds is enough to detect the key of a song, except for those 'striving for innovation' songs. I mean those less tonic pitch used songs.

- Question 3:

  Please design an algorithm that outperforms the two algorithms

- Answer 3:

  I tested several methods to overcome those problems I mentioned above.

Instead of using correlation coefficient as my similarity measurement, I use L2-norm (also known as Euclidean distance), L1-norm (also known as Manhattan distance), and L2-norm, coefficient hybrid method (I called L2-coef) to measure the similarity between chroma vector and Krumhansl-Schmuckler's profile.

For total accuracy, L2-coef method with rate = 0.8 performed the best. We can get higher total accuracy than correlation coefficient.

For 'blues', 'metal' and 'hiphop', L2-norm method highly improved the performance than correlation coefficient.

The table below shows the top accuracy for each method and each genre.

Method	Total Accuracy	pop	Blues	Matel	Hiphop	Rock
Correlation coefficient	45.02%	66.06%	28.67%	45.05%	27.16%	55.91%
L2-norm	45.90%	48.19%	52.44%	57.54%	34.93%	35.00%
L1-norm	10.75%	8.72%	12.95%	14.73%	5.92%	10.71%
L2-coef (rate=0.7)	48.79%	54.68%	47.04%	57.95%	34.93%	47.65%
L2-coef (rate=0.8)	49.20%	59.46%	41.32%	56.23%	32.96%	53.97%
L2-coef (rate=0.9)	49.13%	65.53%	35.81%	53.01%	30.74%	58.26%

The following is the final result of the ‘blues’ genre, which used to be the worst result.