Music Playback Volume Prediction

Zongwei Wang

Abstract: My method is to dig out several features of songs or artists, then use such features to contrast a polynomials regression model with the index equals 1 or 2 to predict the everyday music playback volume. Then calculate the two theta, compare them and choose the better theta to calculate the F.

1 Introduction

After seven years of development and precipitation, Ali Music currently has millions of music library resources, daily tens of millions of users are active on the platform with hundreds of millions of audition, collection and other acts. In the original artists and works, but also has tens of thousands of independent musicians, upload tens of thousands of original works a month, the formation of more than hundreds of thousands of tracks of the original works of art, such a large database of data resources for music trends to grasp having a very important guiding role.

This exercise is based on the history data of Ali music users and expects students to dig out the artists who will become the trend by predicting the amount of auditions of artists in each stage of Ali music platform so as to realize accurately control of popular trends.

2 Related Work

2.1 Polynomials Regression Model

Polynomial regression problems can be solved by transforming variables into multiple linear regression problems.

The model is called a univariate polynomial regression model as follows:

$$y = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

3 Proposed Method

3.1 Processing Datasets

First calculate the singers and songs daily broadcast volume, collection and download number, processing into files artist_p_d_c.txt and song_p_d_c.txt.

In the other sides, we only pick up the songs which has over 200 playbacks and over 20 downloads to improve the result and calculate effectiveness.

3.2 Extract Features

We can extract the following features:

Time difference: songs publish time songs play time

Daily download value

Daily collection value

3.3 Models

We should use degree equals 1 and 2, train two **Polynomials Regression Model**. Then use the trained models to predict two results.

3.4 Theta

$$\sigma_{i,j} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left((S_{i,j,k} - T_{j,k})/(T_{j,k})\right)^2}$$

T is the artists true playbacks, S is the artists prediction.

j is artist, k is the day.

We can calculate two Theta, comparing them and choose the better as the end Theta.

3.5 F

$$F_i = \sum_{jew}(1 - \sigma_{i,j}) * \phi_j$$

among them:

$$\phi_j = \sqrt{\sum_{k=1}^{N} T_{j,k}}$$

3.6 Result

Our result will be 6202.12931728.

4 Conclusion

Using this method we can get a qualified result, but we can do more, here I propose a new idea that we can use the Initial heat, Gender and Language to train every song, get a new value as a new feature. It may be improve our F value.