

Predictions of 59th GRAMMY (2017) Winner and Nominees

Dylan Ong, Eric Hao, Max Schuman, Siyu Zhang

Erichao2018@u.northwestern.edu

Northwestern University, EECS 349 Machine Learning 2016

Our task is not only to predict the winner of the 2017 GRAMMY Record of the Year award, but also to present the songs most likely to be nominated for the award. Because our model can be easily applied to future GRAMMY seasons, these results will yield insights into subtle differences between popular expectations and actual winners, especially in upset years, and can help provide context for identifying shifts and trends in popular music. The output of our project is a rank of songs based on their probabilities to win the GRAMMY Award, given a list of the current top 100 songs eligible for the award and their relevant attributes.

We used a multilayer perceptron model with two hidden layers to make our predictions, which computed the predicted winner score for each song. (In the data, winner_score is 1 if a song won and 0.2 if a song was nominated but did not win). Then, we ranked every song in each year by its winner score to evaluate our predictions. In order to take the fleeting fashion trend into consideration, we assigned each song an associated weight based on its year so that songs from recent years can have more significance when we predicted the winner and nominees for 2017. For each song, we considered 23 attributes, 5 of which are nominal and 19 of which are numeric. 4 nominal attributes are: year, genre, key, mode, time_signature. 19 numeric attributes are: popularity, danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence, tempo, duration_ms, word_count, reading_ease, polarity, subjectivity, followers, listeners, play_count.

We scrapped data from Spotify and Last.fm to obtain various attributes. The data set we have compiled includes roughly 5400 songs from 1958 to 2015. All of these songs were part of the Billboard Year-End Top 100 List. There are 4955 songs from 1958 to 2015 used as training set and 450 songs from 2010 to 2015 used as validation set. We applied the following equation to set up the weight of each data entry based on the year of release.

$$W = \frac{2}{1 + e^{\frac{2015 - [YearOfSong]}{5}}}$$

The validation result is satisfiable.

Year	Position of actual winner in our rank	Number of actual nominees in our rank	Positions of actual nominees in our rank
2011	1	6	1, 4, 5, 7, 35, 78
2012	1	6	1, 14, 16, 22, 53, 64
2013	5	5	5, 7, 9, 29, 32
2014	1	5	1, 19, 53, 65, 71
2015	21	5	1, 10, 13, 21, 41

The result for 2015 is not as accurate as the results for 2011 and 2012 because the model we used to predict the results for 2011-2015 is trained on data from 1958 to 2010, which has smaller gap from 2011 than from 2015. Due to the transience of fashion, the fashion trend in popular songs may have already changed, leading to the decrease in accuracy of our validation.

However, this will not be a problem when we make our predictions for 2017 since the model we use is trained on data from 1958 to 2015, with songs from recent years being assigned more weights thus exerting more influence.

Before the multilayer perceptron model with two hidden layers, we tried Logistic Regression Model to judge the probability of a song being classified as winner and using the probability to rank songs and used the same method to get probability of a song being nominated. We did predictions for the winner and nominees separately and ignored the correlation between them. Expectedly, our result for Logistic Regression Model was not satisfying. Using a logistic regression model with 20-fold cross-validation we found that 66 of the 286 nominees in our data set are correctly classified as nominees and 7 of the 58 winners are correctly classified as winners.

However, classification recall is a flawed metric for our purposes, as the nature of our task involves ranking songs within years to determine the worthiest Record of the Year winner in a given year, not identifying the winners from all years in a batch of songs released over the course of sixty years. Therefore, we incorporated information about songs that were nominated but didn't win and tried to predict the winner_score. For training data, we assigned the winner_score of a song to be 1 if it won and 0.2 if it was nominated but did not win. In other words, besides the year of release, we also assigned weight to a song based on whether it won or was nominated. And then we tried the multilayer perceptron model, which has better result (listed as above).

To make our predictions for 2017 GRAMMY Winner and Nominees, we first collected the candidates for 2017 Grammy Award. As required, songs that are eligible for the 59th (2017) GRAMMY must be released between October 1st 2015 and September 30th 2016. Since there are still several months before the deadline and we cannot predict which songs will be released before September 30th 2016, we only considered songs released from October 1st 2015 to present. We chose our candidate songs from the top 100 billboard list for each week since October 1st 2015. However, this by no means guarantees that the songs were released after October 1st 2015. So there may be some older 2015 songs that are not eligible. We took the liberty of killing any song that was before October 2015. In total, there are 286 eligible candidates for 2017 Grammy Award with 23 attributes for each.

Then, we made our predictions based on the multilayer perceptron model we trained on data from 1958 to 2015. The result is as follows. The form contains top 30. More results are on the website.

Rank	Song	Artist	Winner_Score
1	Ultralight Beam	Kanye West	0.248
2	Sandcastles	Beyonce	0.151
3	Blackstar	David Bowie	0.113
4	Hold Up	Beyonce	0.083
5	Water Under The Bridge	Adele	0.076
6	Lazarus	David Bowie	0.049
7	Love Yourself	Justin Bieber	0.047
8	Zero	Chris Brown	0.046
9	Tennessee Whiskey	Chris Stapleton	0.044
10	Traveller	Chris Stapleton	0.044
11	Photograph	Ed Sheeran	0.035
12	Purpose	Justin Bieber	0.033
13	Dibs	Kelsea Ballerini	0.032
14	Die A Happy Man	Thomas Rhett	0.031
15	7 Years	Lukas Graham	0.030
16	All I Ask	Adele	0.029
17	Mark My Words	Justin Bieber	0.025
18	Life Is Worth Living	Justin Bieber	0.024
19	Lonesome Broken And Blue	Adam Wakefield	0.024
20	Blessings	Chance The Rapper	0.023
21	Lost Boy	Ruth B	0.023
22	God Only Knows	Jordan Smith & Adam Levine	0.022
23	Nobody To Blame	Chris Stapleton	0.022
24	Moolah	Young Greatness	0.020
25	Cheerleader	OMI	0.020
26	All The Way Up	Fat Joe & Remy Ma Featuring French Montana & Infared	0.018
27	Gonna Know We Were Here	Jason Aldean	0.017
28	Stay A Little Longer	Brothers Osborne	0.017
29	Huntin' Fishin' & Lovin' Every Day	Luke Bryan	0.017
30	Gonna	Blake Shelton	0.017

From the prediction result, we can see that Top 5 are far ahead others. And there are subtle differences from No.5 to No.10, No.11 to No.16, and No.17 to No.30. So, we have much confident on Kanye West and Beyoncé this year. For Justin Bieber, Beyoncé, Adele and Chris Stapleton, they have several songs on Top 30, which increases their possibility to win the GRAMMY Awards.

We are looking forward to the 59th GRAMMY in next year to see how accurate our prediction is. For

future suggestions, there are still some details can be improved. First, we only use songs between October 1st 2015 and June 1st 2016 as our candidate songs. However, there may be some more popular songs going to be released in the next several months. So, if possible, songs that will be released should be added into our candidate dataset. Second, the dataset we used to train our model is extracted from Spotify and last.fm. To control the size of our dataset to a moderate number, we only picked up songs Billboard Year-End Top 100 list. But it is possible that songs not in the list were also nominated before. Thus, if possible, we can increase the size of our training dataset to include more popular songs.

Teamwork:

Dataset: Dylan Ong, Eric Hao, Max Schuman

Training models: Max Schuman, Siyu Zhang

Report: Siyu Zhang

Website: All