G B

ology@github

Search this website .

Predicting Beatles Song Authorship with scikit-Learn

2018-08-13 BY GENE

Yesterday I read about a 10 year effort to predict the author of Beatles songs by lyrics, tonal contour and chord analysis.

Hmm very curious. Can I replicate a part of this I wonder? tl;dr: beatles.py

I found a table listing all the songs and who wrote them (TXT). Also I knew I could write a program to harvest all the lyrics. So that's what I did! I grabbed the author table and all the lyrics, and put them into simple text files (ZIP). This in hopes that I could use machine learning to divine the author from just the lyrics. Fingers crossed!

First off I import a few things from scikit-Learn that will be used at the end of the program.

Next I collect the songs, their authors and their lyrics into a dictionary. In the midst of this logic, I constrain the list of songs to only those written by *either* John Lennon or Paul McCartney. This simplifies our problem to just two authors.

With the dictionary in hand, I then create the **X** and **y** lists needed to feed to the learning algorithms. **X** is the list of lyrics. The **y** list is the authors.

Since this is a text classification problem, I will first use the CountVectorizer to convert the raw text into word counts. Next I use MultinomialNB to learn the lyrics-author association.

I don't expect a great prediction accuracy, partly because I am a pessimist at heart… And sure enough, the accuracy turns out to be 0.625 or 62.5%. Hrm.

FILED UNDER: DATA, SOFTWARE
TAGGED WITH: BEATLES, MACHINE LEARNING, PYTHON

Epistemologist-at-large

^ **Top**