

Predicting Song Hotness

Problem statement

If I'm going to be a mess, I may as well be a hot mess. – Mindy Kaling

The days of buying vinyl records and CDs is all but over. Most music consumers now purchase downloadable songs on online retailers such as amazon.com, or purchase subscriptions to services such as Spotify, Pandora, or iTunes. Record and CD sales used to be a measure of a song's popularity, but now we need to rely on download rates and other similar measures to assess the hotness of music.

The *Million Song Dataset*, a project developed from a grant from the Nation Science Foundation, consists of audio features and metadata for a million popular songs. One of the variables available in the data set is a song's hotness, a variable constructed by The Echo Nest (which is now owned by Spotify). The goal of this project is to predict a song's hotness from other features of a song. Features that may be particularly predictive include danceability, energy, tempo, and year of the song, though your job is to explore the many features that may be predictive of song hotness.

Key Challenges: Wrangling data, visualization, statistical modeling, machine learning.

Data resources

You should plan to work with the 10,000 song subset in developing your prediction approach (see the link below). All of the information you need to know about the data set are on the Million Song Data set main page.

- Home page for Million Song Data set:
<https://labrosa.ee.columbia.edu/millionsong/>
- Link to 10,000 song subset:
<https://labrosa.ee.columbia.edu/millionsong/pages/getting-dataset#subset>
- Annotated list of data fields:
<https://labrosa.ee.columbia.edu/millionsong/pages/example-track-description>
- Example project on recommendation systems derived from the Million Song Dataset:
<http://www-personal.umich.edu/~yjli/content/projectreport.pdf>

References

1. Bertin-Mahieux, T., Ellis, D. P., Whitman, B., & Lamere, P. (2011). The million song dataset.

2. McFee, B., Bertin-Mahieux, T., Ellis, D. P., & Lanckriet, G. R. (2012, April). The million song dataset challenge. In Proceedings of the 21st International Conference on World Wide Web (pp. 909-916). ACM.
3. Schindler, A., Mayer, R., & Rauber, A. (2012, October). Facilitating Comprehensive Benchmarking Experiments on the Million Song Dataset. In ISMIR (pp. 469-474).
4. Hu, Y., & Ogihara, M. (2012, August). Genre classification for million song dataset using confidence-based classifiers combination. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (pp. 1083-1084). ACM.