
Modeling Musical Influence with Topic Models

Uri Shalit

URI.SHALIT@MAIL.HUJI.AC.IL

ICNC-ELSC & Computer Science Department, The Hebrew University of Jerusalem, 91904 Jerusalem Israel

Daphna Weinshall

DAPHNA@CS.HUJI.AC.IL

Computer Science Department, The Hebrew University of Jerusalem, 91904 Jerusalem Israel

Gal Chechik

GAL.CHECHIK@BIU.AC.IL

The Gonda Brain Research Center, Bar Ilan University, 52900 Ramat-Gan, Israel

1. Ablation study

In our work we used three sets of features. Two sets are sets of local features, pertaining to a short audio segment:

- **chroma**: A 12-component vector corresponding to the 12 pitch classes C, C#, D to B, with values ranging from 0 to 1 that describe the relative dominance of every pitch in the chromatic scale.
- **timbre + max. loudness**: A 13-component vector describing the quality of a musical note or sound that distinguishes different types of musical instruments, or voices. The first 12 components are derived using the 12 top PCA components of a descriptor similar in nature to MFCC (Mermelstein, 1976). The last corresponds to the peak dB value of the segment.
- **global features**: The third is a set of global features - tempo, key, mode and time signature (meter). These describe overall aspects of a song.

In order to assess the relative contribution of each feature type, we conducted an ablation study. We trained the model using one set of features at a time, and evaluated the mean Spearman correlation with `allmusic.com`'s influence rank (described in Section 5.3 of the main text). For the timbre and chroma features, the bag-of-words representation was built in the same manner explained in the main text, by quantizing the per-segment descriptors. We used the same value for the K-means, $K = 5000$. The global features are 33 separate words describing each song, as described in section 4 of the main text.

The results are presented in Figure 1. Using all 3 sets of features gives a mean Spearman correlation of 0.15; using only timbral features gives a correlation of 0.12; using only chroma features or only global features yield

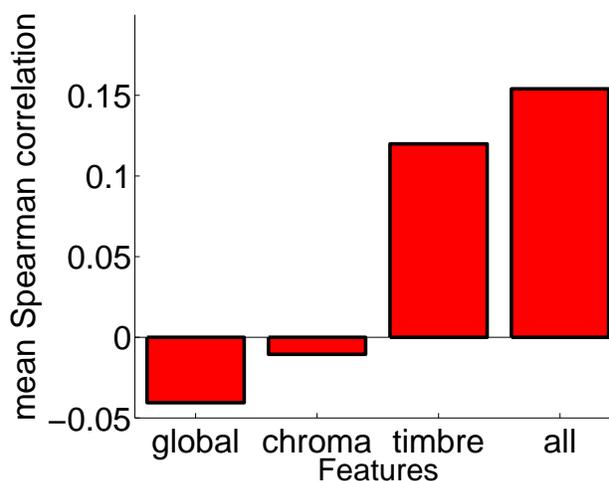


Figure 1. Mean per-epoch Spearman corr. with `allmusic.com`'s influence rank for a 10-topic model trained with different features: Only timbre and loudness features, only chroma features, only global features and finally a model using all features in conjunction.

non-significant mean correlations of -0.01 and -0.04 , respectively. It seems that on their own, only the timbral features enable building a strong model. However, there is a synergy effect in place, and adding all three feature sets together gives a better overall correlation. These results are broadly in line with the work by (Serrà et al., 2012). They show that the use of timbral features is the main differentiating factor between songs across the years, while chroma features are much more stable. In our terms, we might say that this is the aspect most given to evolution and influence between songs over time.

References

- Mermelstein, P. Distance measures for speech recognition, psychological and instrumental. *Pattern recognition and artificial intelligence*, 116:91–103, 1976.
- Serrà, J., Corral, Á., Boguñá, M., Haro, M., and Arcos, J.L. Measuring the evolution of contemporary western popular music. *Scientific Reports*, 2, 2012.