Music Lyric Topic Modelling

Related Work

Music genre classification based on lyrics

View of Data Science Approach to Compare the Lyrics of Popular Music Artists (uii.ac.id)

- Not on genres but on popular artists
- 10% accuracy instead of 1% (random)
- Use of classification: See whether classes can be well separated

aclanthology.org/C14-1059.pdf

- Cites background evidence: lyrics important in genre classification
- Classify (SVM) genre based on n-grams and other features such as typ-token ratio
- Compare to human perception
- Genres looked at: Blues, Rap, Metal, Folk, R&B, Reggae

Performance of topic models for topic annotation of music lyrics

Assessing Quality of Unsupervised Topics in Song Lyrics | SpringerLink

- How well are topics from topic models compared to manual topic annotations?
- Unfortunately cant read the paper (fuck SpringerLink)

lyric_topic_modeling.pdf (alenlukic.com)

- Like the paper above just with additional lyrics and additionally using the Pachinko Allocation (PAM)
- Qualitative Evaluation: Compare to supervised topics, look whether there are some without any similarity
- Quantitative Evaluation: Similarity (normalized maximum similarity, Kurtosis) with the manual topics (using same dataset as paper above + using stanford topic modeling toolbox to get supervised labels)
- Conclusion: More data proofs useful and PAM finds supertopics

Indonesians' Song Lyrics Topic Modelling Using Latent Dirichlet Allocation | IEEE Conference Publication | IEEE Xplore

- Indonesian Lyrics are a challenge because of their high poetic level
- Very method-oriented preprocessing and LDA:
- Also evaluated each topic manually. Comparison to audiosparx.com to name topics
- Automatic Evaluation of perplexity

ismir-2014-sasaki.pdf (kyoto-u.ac.jp)

- Cited very often
- Gives an interface to search for songs by topics instead of name or artist
- Also uses LDA

Sentiment analysis based on topics of lyrics

Exploiting Topic Modelling to Classify Sentiment from Lyrics | SpringerLink

- Fuck SpringerLink
- Tests LDA and Heuristic Dirichlet Process (HDP)

Music mood classification based on lyrical analysis of Hindi songs using Latent Dirichlet Allocation | IEEE Conference Publication | IEEE Xplore

- Analyze lyrics to detect mood of listener
- Uses LDA
- Uses own data (manually annotated): distinguish sad and happy
- Dont get when they actually classify? Chose 10 topics → easily distinguishable between two classes → can get probability distribution of each topic in each song
- Evaluation terrible: Just look how many are in each class not even accuracy like wtf

Other uses of topic models for music lyrics

nips2013tm_submission_25.pdf (cornell.edu)

- Researches rap music over time and place, focusing on shift from “ghetto” to “hood” (rappers conception of place)
- Rap lyrics have been studied before: First time using big database
- Uses LDA for initial topic model, then Dirichlet-multinomial regression (DMR) to generate topics conditioned on times, places and artists
- Evaluation using posterior predictive checking

Topic Models for Classification:

Statistical topic models for multi-label document classification | Machine Learning (springer.com)

- Essentially: Need label per word not per document: Can be sampled
- Learning is not that trivial. If we have to implement that ourselves we cannot do this
- Also: its for multi-label annotation, which may not be the case for us

Using LDA Topic Models as a Classification Model Input | by Marc Kelechava | Towards Data Science

- Uses topic distribution as feature for a classifier
- Would be interesting for us: What are the resulting topics?
    - That would be the DH part: Making sure that it makes sense
    - Also decide on how many topics and so on, just see whats happening
- Says paper: fp224-phan.dvi (sourceforge.net) does similar stuff, but didn’t read it

Ideas

- Could use topic modeling to classify genres
    - Do that similarly to mood classification
    - Compare to methods without using topic models