TRACKING THE EVOLUTION OF UNDERGROUND MUSIC CULTURE VIA DEEP SENTIMENT ANALYSIS

Christian Brown (csbrown@stanford.edu)

DEPARTMENT OF ENERGY RESOURCES ENGINEERING, STANFORD UNIVERSITY

Motivation

Resident Advisor (RA) music reviews have been describing and critiquing the latest track releases in the underground music scene since 2001. By understanding which words or phrases are given more weight when predicting a reviews sentiment, and whether these words or phrases are valuable for predicting timeframe (era) of publication, interpretations can be made on how the underground music community's preferences for specific musical elements changed over the last 20 years.

Objectives

- Develop and train a pair of models that classify a review as either positive or negative, and the era of release
- Identify the keywords in the review that the model used to make its predictions
- Compare the keywords used for sentiment classification and timeframe of release in order to identify how musical qualities of underground music have changed over time

Method – step by step

1. Train and develop a pair of models with identical architecture for sentiment and era classification
   - Single-Layer LSTM with dropout, Adam optimizer
   - Loss:
   $$\mathcal{L} = \frac{1}{m}\sum_{i=1}^{m} y_i \log(\hat{y}_i) + \frac{\lambda}{m}\sum_{k=1}^{4}\|W\|^2$$
2. For each review, assess model activations and visualize similarities of activations for time--steps via cosine similarity:

   For t = 1 .. T_x-1:
   $s_t = s(a_t, a_{T_x})$

   Where s_t given by $s(a_1, a_2) = \cos(\theta_{a_1,a_2}) = \frac{a_1 \cdot a_2}{\|a_1\|\,\|a_2\|}$
3. Obtain keywords via selecting words that result in large changes in similarity prior to convergence
4. Count over all reviews the number of instances a keyword is observed for each class (sentiments, timeframes)
5. Identify which common keywords are unique to each class, compare instances of keywords in sentiment classification and timeframe classification

Data

- ~17,000 RA music reviews from 2001-2017 via Kaggle[1]
  - *Rating*, *Year*, Genre, Label…
- Preprocessing- Remove review features, non-alphabetical characters, stop-words, etc.
- Train/Dev/Test 70/15/15

… with earthy drums leading to a rollicking bassline and classic Detroitesque piano stabs. The melodic techno of the first half gives way to a darker second section after the breakdown, pushed along by bolder percussion and rhythm that finish the track with a flurry …

Fig 1. Review excerpt for 'Heiligendamm' by Kollektive Turmstrasse (2008)

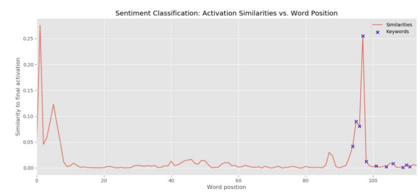
Label	Feature used	Class Definitions *(even across classes)*
Sentiment	Rating	i. Negative: < 3.7, ii. Positive: > 3.7
Era	Year of Release	i. 2001-2009, ii. 2010-2013, iii. 2014-2017

Table 1. Description of labels and definitions used for training classifiers.

Results

Classifier	Dev-set Accuracy	Dev-set Recall	Dev-set Precision
Sentiment *(2-class)*	66.0%	78.0%	64.9%
Era *(3-class)*	80.1%	78.5%	86.4%

Table 2. Performance metrics on development set of optimized classifiers



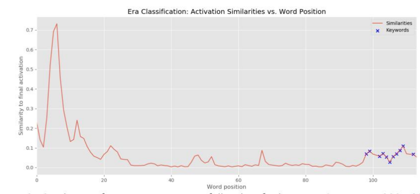


Fig 2. Similarity of activations successfully classified review (negative, 2001-2009) review for 'Heiligendamm' by Kollektive Turmstrasse

Keyword Analysis

- For each class (five total), keyword instances were counted
- Top keywords for each class were selected by returning the n_s or n_e most frequent words
- Keywords were considered unique to a class if they were not listed in the most frequent keywords for other classes

	Positive	Negative	Neither
2001-2009	Solid	Smooth, flip, nice, breaks	Original, massive
2010-2013	Vocal, touches, tight, synth, drum, groove, pop	n/a	Kind
2014-2017	Percussion, kicks, glowing, hard	n/a	Atmosphere, package

Table 3. Subset of unique words for each subclasses (n_s = 200, n_e = 50)

- Words associated with positive reviews tended to change overtime
- Lack of negative words acting as unique identifiers of era indicate the vocabulary of negative sentiment hasn't changed substantially

Conclusions

- Higher accuracy of era classifier indicate language used in reviews continues to evolve over time
- Magnitude of neuron activation similarities are not necessarily the same across classifiers, indicating certain words are good indicators of sentiment, but not era and vice versa
- Positive language evolves over time, whilst negative words appear relatively consistent

Future Work

- Explore model with attention to handle longer review sequences
- Data augmentation and additional sub-setting to obtain activations specific to era/sentiment subclasses
- Extend definition of keywords to capture points that solicit stronger responses in similarity graphs
- Use less naïve approach to identify keywords (phrase extraction)

References

1. https://www.kaggle.com/marcschroeder/17-years-of-resident-advisor-reviews