EXPLORING AUDIO AD EFFECTIVENESS

MATTHEW PROCKUP, PUYA - HOSSEIN VAHABI, SAMANEH EBRAHIMI

CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS
ML4AUDIO WORKSHOP 2017
PERSONALIZED MUSIC ENJOYMENT AND DISCOVERY
What is Pandora?

known for radio

on-demand too!
Pandora’s Team

**Scientists and Engineers:** Music Information Retrieval, Recommender Systems, Listener Engagement, Library Content Management, Advertising.

**Curators/Librarians:** Content management, discovery, curation and exploration.

**Music Analysts:** Experts label tracks on >500 music attributes.
Pandora’s Scale

300M+ registered users

~73M monthly actives in the United States

#1 in mobile time spent. More than Spotify, Facebook, YouTube, Snapchat and Instagram. (comScore)

~23 hours/month, ~10% all U.S. radio listening

12B+ stations, 85B+ thumbs
Advertisement at Pandora

Audio and Video Advertisements

Pandora is the largest publisher of digital audio advertising in the US

Target at exactly the right time and in the right context (2000+ targeting segments)

Pandora is #1 in 84 US radio markets

$1.4B Revenue per Year
AUDIO ADS
Are they engaging?
Introduction: Audio Ad Experience
Which ad is the most engaging?
Why is important?

Ad Ranking
Bid * Relevance
Better UX
More engaging ads for each of you.

What creates an engaging ad?
Problem Definition

Let $A = a_1, a_2, \ldots, a_N$ be a set of ads, our objective is to learn a function $f : A \rightarrow R^{\geq 0}$ that indicates the ad engagement.

Open problem

How do we measure the audio ads engagement?
Related Ad Experiences
Sponsored Search, Display Ads, etc.

Audio Ads can’t be clicked
Clicks on companion audio banner!
Related Ad Experiences

Sponsored Search, Display Ads, etc.

Clicks are not an engagement metric!
Measure Engagement

Exploit **Dwell-time** on landing page from **clicking** on the add banner.
Dwell-Time as a metric of engagement
The Long Click

A **long click** combines both an click and a follow-up **engagement**

### Long Click Rate (LCR)

\[
LCR = \frac{\text{long clicks}}{\text{impressions}}
\]

### Robust-LCR (R-LCR)

\[
R-LCR = \frac{\text{users long clicked}}{\text{users}}
\]

30%  
Bad Ads  

30%  
Good Ads
Advertisement Components

- Recording Quality
- Tone & Tempo
- Brand Trust
- Message
- Speakers
- Music & SFX
Advertisement Components

- Recording Quality
- Tone & Tempo
- Brand Trust
- Message
- Speakers
- Music & SFX
ACOUSTIC FEATURES
informed by speech and music-IR
Acoustic Features

**Timbre**

Temporal Features and Dynamics (rms, zero-crossing, etc)

Mel-Frequency Cepstral Coefficients (MFCC)

Delta MFCCs

Mel Spectral Patterns (MSP)

- Gender
- Multiple Speakers
- Music vs. Speech
- SFX

Music vs. Speech
Acoustic Features

Rhythm and Timing

Tempogram
Tempo/Pulse Estimate
Relative Tempogram Ratios
Mellin Scale Transform

Event Timing
Signal Repetition
Pulse & Pace
Background Rhythm
Acoustic Features

Harmonic Organization

Seemingly tangential, but ...

Chroma

Key Template Correlations

Chord Template Correlations

Chord Estimate Histograms

Monotone vs. Animated

Clear Pitched vs. Raspy/Breathy

Background Harmony
MACHINE LEARNING
modeling advertisement quality
Ad Quality Modeling
interpretable models

**Single**
- Logistic Regression (LR)
- LR with L1 Regularization for Feature Selection (LR+L1)
- Support Vector Machine (SVM)

**Ensemble**
- Random Forest (RF)
- Gradient Boosted Decision Trees (GBT)
Ad Quality Modeling

experimental setup

**Dataset**

- Ads for May-June 2017
- 90%/10% train/test split
- 128kbps mp3 files and acoustic features
- 9k LCR examples
- 7k R-LCR examples

**Experiments**

- 10-fold stratified cross-validation
- Predict LRC, R-LCR
- Report AUC
RESULTS
Quality Prediction Results

weighted logistic regression wins here

![Chart showing AU-ROC for different models: LR, LR+L1, SVM, GBT, RF.]
Feature Importances

LCR Top Feature Correlations (224 selected)

R-LCR Top Feature Correlations (237 selected)
Feature Correlation Interpretation

**MFCC**: (-) noisy shape. (+) basic shape clearly variant formants

**DMFCC**: (+) (-) shape change. Speech should be clear and variant.

**MSP**: (-) high variance in high and low mel frequencies, a noisy environment. (+) stability in middle frequencies, i.e. speech

**Tempo (TG, TGR)**: (-) corr. with fast pulse repetition. (+) corr. slower pulse repetition.

**Harmony Features (SICK, SICH)**: basically a wrapped spectrum (+) moderately animated pitch shifts. (-) too much shift or too little shift, suggesting too much or to little speech animation.
Qualitative Analysis

exploring clusters that tend to have high or low quality

Test Set Audio Features → K-Means
Some Interpretations

calls to action

**High Quality**
- clear mid-paced voice
- solo speaker
- moderate expression, conversational
- proper background balance
- little SFX

**Poor Quality**
- Fast talking
- Long-winded explanations
- Jarring background and SFX
- Bad mix of background and SFX
- Low quality recordings
GOING (A LITTLE) DEEPER
Neural Networks
multi-layer perceptron (MLP)

Audio File (mp3) → Acoustic Feature Extraction → 2440 features → Fully Connected Layers
- Relu activation
- Drop out

Hidden Layer I → Hidden Layer II → Output Layer (Ad Quality)
- Sigmoid
Deep Learning
convolutional neural network

Hidden Layers

Audio File (mp3) → Audio Spectrogram → Convolutional layers, Maxpooling, Global pooling → Dense Layers (ReLU activation, Dropout) → Output Layer (Ad Quality) Softmax
Quality Prediction Results

Deep Networks are More Effective

AU-ROC

LR, LR+L1, SVM, GBT, RF, MLP-2, MLP-3, CNN-4

LCR, R-LCR
Conclusion

Ad engagement prediction is important for ranking

Engaging Ads
Speak slowly, slightly animated, and with clear pronunciation. Music is fine if not jarring.

Results
AUC ~0.8 for engagement prediction using deep learning Feature based machine learning approach is at least 10% less precise
THANK YOU - Q&A

“Predicting Audio Advertisement Quality”
Samaneh Ebrahimi, Puya Vahabi, Matthew Prockup, Oriol Nieto
ACM Web Search and Data Mining (WSDM) 2018