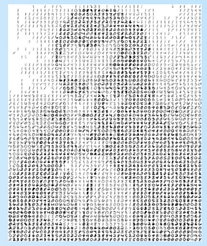
Applying Psychological Models of Generalization to Music Recommendation Systems

Tiffany Hwu
tiffanyhwu@berkeley.edu

Joshua Abbott
joshua.abbott@berkeley.edu

Thomas Griffiths
tom_griffiths@berkeley.edu



Introduction

In generalization, stimuli are defined within a psychological space, in which previously experienced stimuli are used to create generalizations about newly presented stimuli (Shepard 1987). Similarly, a person who is recommending songs will use their prior musical knowledge or intuitions to find songs of similar taste.

The main objective of this work is to evaluate the effectiveness of applying psychological models to large -scale online datasets which contain the listening histories of users.

Background

Applying the Bayesian Framework of Generalization (Tenenbaum & Griffiths 2001)

The Bayesian generalization framework has been successfully used in a variety of different domains, and consists of the following parts:

Hypothesis space of positively seen examples
Ex. Playlists created by users
Prior – Erlang distribution
Intermediate size playlists are more informative
Likelihood – size principle
Smaller playlists more focused than larger playlists, given the prior

$$P(X|h) = \begin{cases} 1/|h|^n & x^{(i)} \in h \\ 0 & otherwise \end{cases}$$

Bayes Rule

$$P(h|X) = \frac{P(X|h)P(h)}{\sum_{h' \in H} P(X|h')P(h')}$$

Probability of generalization

$$P(y \in C|X) = \sum_{h \in H} P(y \in C|h)p(h|X)$$

Sample input and output
X={'Baby One More Time, Womanizer, Toxic'}
Y={'Me Against the Music', 'Crazy','Oops I Did It Again'…}

Methods

Datasets: Million Song Dataset, Art of the Mix 2011 dataset
Split datasets into test and train binary matrices

Constructing a Hypothesis Space

Songs	h1	h2	h3	h4	h5	h6
s1	1	0	1	0	0	1
s2	0	0	1	0	0	0
s3	0	0	0	1	0	1
s4	0	1	0	0	0	1
s5	1	0	0	1	0	1
s6	1	0	0	0	1	0

Psychological Models	Traditional Models
Bayesian Generalization	TF-IDF
Prototype	Bayesian Sets
Exemplar	Popularity

Metrics

Mean Average Precision (mAP)
Discounted Cumulative Gain (DCG)
Mean Reciprocal Rank (MRR)
Precision at 10 (P at 10)

Evaluation Conditions

Popularity Threshold
Query Size
Qualitative Review

Results

X={'Baby One More Time'}

Bayesian Generalization	TF-IDF	Bayesian Sets
*Crazy	*Crazy	Where Is My Mind?
*Sometimes	*Sometimes	In the Aeroplane Over the Sea
*Lucky	*Lucky	Pink Moon
Genie in a Bottle	My Heart Will Go On	Love Will Tear Us Apart
Larger Than Life	Wannabe	Such Great Heights
Leave	Larger Than Life	Asleep
Livin' La Vida Loca	Genie in a Bottle	Just Like Heaven
*Stronger	Livin' La Vida Loca	Everlong
My Heart Will Go On	Bye Bye Bye	Hallelujah
Bye Bye Bye	The Bad Touch	Fake Plastic Trees

X={'Baby One More Time', 'Womanizer', 'Toxic'}

Bayesian Generalization	TF-IDF	Bayesian Sets
*Me Against the Music	99 Problems	Yeah!
*Crazy	If You Could Only See	*Crazy
*Oops I Did It Again	Cold Hard B****	*Sometimes
*Everytime	Last Nite	Confessions Part 2
*Sometimes	And She Was	*Everytime
*Lucky	Sweet Child O' Mine	*Me Against the Music
*I'm a Slave 4 U	Gent	*Oops I Did It Again
*Stronger	Take the Skinheads Bowling	Burn
*Boys	Could You Be Loved	Come Clean
*Don't Let Me Be the Last to Know	Just a Phase	*Lucky

We see that Bayesian Generalization is good at discriminating between artist and genre, depending on the size of query.
For Michael Jackson songs, Bayesian Generalization recommends MJ songs if we provide 3 MJ songs in the query, and a mix of Halloween and MJ songs if we just provide 'Thriller'.

X={'Thriller','Billie Jean','The Way You Make Me Feel'}

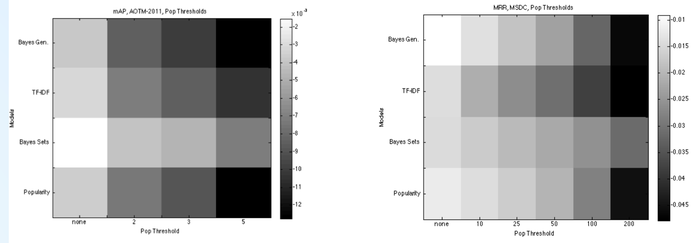
Bayesian Generalization	TF-IDF	Bayesian Sets
*Bad	*Smooth Criminal	The Monster Mash
*I Just Can't Stop Loving You	Tiny Dancer	*Smooth Criminal
*Smooth Criminal	Like A Prayer	Nightmare on My Street
*Man in the Mirror	*Let's Get It On	*Bad
*Wanna Be Startin' Somethin'	I Believe in a Thing Called Love	*Can You Feel It
*PYT	*Money	Love is a Battlefield
*You Rock My World	Kiss	Every Day is Halloween
*Baby Be Mine	*Bad	*I Just Can't Stop Loving You
*Why You Wanna Trip on Me?	Dead Man's Party	Shake Your Groove Thing
*Speechless	Halloween	*Stranger in Moscow

X={'Thriller'}

Bayesian Generalization
Halloween
The Monster Mash
Werewolves of London
I Put a Spell on You
*Billie Jean
Dead Man's Party
Girls Just Wanna Have Fun
*Smooth Criminal
Ghost Town
*The Way You Make Me Feel
I'm a Mummy
Video Killed the Radio Star

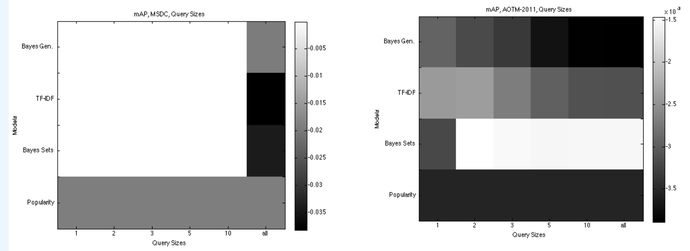
Results Pt. 2

mAP by Popularity Threshold



For both MSD and AOTM-2011 datasets, performance increases as popularity threshold increases. Bayesian generalization performs better on the AOTM-2011 and TF-IDF performs better on the MSD.

mAP by Query Size



As query size increases, mAP scores generally increase. The MSD does not perform well for any of the query sizes except for 'all'.

Discussion

Differences of results in the MSD and AOTM-2011 datasets might be explained by the fact that AOTM-2011 was created from user-selected playlists while the MSD contained entire listening histories.

We have seen that applying psychological models of generalization can contribute significantly to current systems and perhaps provide more insight into how a human would recommend songs versus how a typical machine learning algorithm would recommend songs.

References:
Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. Science, 237, 1317-1323.

Tenenbaum, J., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. Behavioral and Brain Sciences, 24, 629-641.