Recommending and Searching

Research @ Spotify

Mounia Lalmas

Chalmers University of Technology, 4-5 March 2019
Making AI works at Spotify

Evaluation offline and online

Algorithm(s)

Training & Datasets

Interaction & feedbacks data

Features (item)

Features (user)

Features (context)

Metric(s)

Qualitative research

Business metrics
What we do at Spotify
Spotify’s mission is to unlock the potential of human creativity — by giving a million creative artists the opportunity to live off their art and billions of fans the opportunity to enjoy and be inspired by it.
Our team mission:

Match fans and artists in a personal and relevant way.
What does it mean to match fans and artists in a personal and relevant way?
What does it mean to match fans and artists in a personal and relevant way?
“We conclude that information retrieval and information filtering are indeed two sides of the same coin. They work together to help people get the information needed to perform their tasks.”

“We can conclude that **recommender systems** and **search** are also two sides of the same coin **at Spotify**. They work together to help fans get **the music they will enjoy listening**”.

is this the case?
Home ... the push paradigm
Home

Home is the default screen of the mobile app for all Spotify users worldwide.

It surfaces the best of what Spotify has to offer, for every situation, personalized playlists, new releases, old favorites, and undiscovered gems.

Help users find something they are going to enjoy listening to, quickly.
BaRT: Machine learning algorithm for Spotify Home

BaRT (Bandits for Recommendations as Treatments)

How to rank playlists (cards) in each shelf first, and then how to rank the shelves?
BaRT: Multi-armed bandit algorithm for Spotify Home

Explore vs Exploit

Flip a coin with given probability of tail
If head, pick best card in M according to predicted reward $r \rightarrow$ EXPLOIT
If tail, pick card from M at random $\rightarrow$ EXPLORE
Success is captured by the reward function

Success is when user streams the playlist for at least 30s.

\[ R(t) = \begin{cases} 0 & \text{if } t < 30 \text{ sec} \\ 1 & \text{if } t \geq 30 \text{ sec} \end{cases} \]
Is success the same for all playlists?

Consumption time of a sleep playlist is longer than average playlist consumption time. Jazz listeners consume Jazz and other playlists for longer period than average users.
Personalizing the reward function for BaRT

**Global distribution**

\[
\hat{r}(s)
\]

One reward function for all users and all playlists

Success independent of user and playlist

**Individual distribution**

\[
\hat{r}_{u,y}(s)
\]

One reward function per user \times playlist

Success depends on user and playlist

Too granular, sparse, noisy, costly to generate & maintain

**Co-cluster distribution**

\[
\hat{r}_{u,y}(s)
\]

One reward function per group of users \times playlists

Success depends on group of users listening to group of playlists
Co-clustering


Caveat no theoretical foundation for selecting the number of co-clusters apriori

\[
\text{group} = \text{cluster} \\
\text{group of user x playlist} = \text{co-cluster}
\]
Co-clustering for Spotify Home

Any (interaction) signal can be used to generate the co-clusters.
Reward function per co-cluster using distribution of streaming time
Experiments

Counterfactual methodology, which works like offline A/B

Thresholding methods: mean, additive, continuous & random

Baseline (one threshold), playlists only, users only, both

One week of random sample of 800K+ users, 900K+ playlists, 8M user-playlist interactions

Expected stream rate

Conclusions

Accounting for user experience and playlist consumption matters.

Co-clustering users and playlists surfaces patterns of user experience x playlist consumption.

Using one interaction signal and a simple thresholding method can already provide effective personalised success metrics.
Recommendation in a 2-sided Marketplace

Multiple objective functions

\[ a_t = \arg \min_{\pi} \pi \]

\[ a_t = \arg \min_{\pi} f(\pi_1, \pi_2, \pi_3, \ldots) \]
Recommendation in a 2-sided Marketplace

- Policy I: Optimizing Relevance
- Policy II: Optimizing Fairness
- Policy III: Probabilistic Policy
- Policy IV: Trade-off Relevance & Fairness
- Policy V: Guaranteed Relevance
- Policy VI: Adaptive Policy I
- Policy VI: Adaptive Policy II


ML Lab

An offline evaluation framework to launch, evaluate and archive machine learning studies, ensuring reproducibility and allowing sharing across teams.
Search ... pull & push paradigms
Searching for music

**Large catalog**
40M+ songs, 3B+ playlists
2K+ microgenres

**Many languages**
79 markets

**Different modalities**
Typed, voice

**Heterogeneous content**
Music, podcast

**Various granularities**
Song, artist, playlist

**Various goals**
Focus, discover, lean-back, mood, activity
Overview of the user journey in search

INTENT
What the user wants to do

MINDSET
How the user thinks about results

TYPE/TALK
User communicates with us

CONSIDER
User evaluates what we show them

DECIDE
User ends the search session
<table>
<thead>
<tr>
<th><strong>FOCUSED</strong></th>
<th>One specific thing in mind</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Find it or not</td>
<td></td>
</tr>
<tr>
<td>● Quickest/easiest path to results is important</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>OPEN</strong></th>
<th>A seed of an idea in mind</th>
</tr>
</thead>
<tbody>
<tr>
<td>● From nothing good enough, good enough to better than good enough</td>
<td></td>
</tr>
<tr>
<td>● Willing to try things out</td>
<td></td>
</tr>
<tr>
<td>● But still want to fulfil their intent</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>EXPLORATORY</strong></th>
<th>A path to explore</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Difficult for users to assess how it went</td>
<td></td>
</tr>
<tr>
<td>● May be able to answer in relative terms</td>
<td></td>
</tr>
<tr>
<td>● Users expect to be active when in an exploratory mindset</td>
<td></td>
</tr>
<tr>
<td>● Effort is expected</td>
<td></td>
</tr>
</tbody>
</table>

Focused mindset

65% of searches were focused.

When users search with a Focused Mindset

Put **MORE effort** in search.

Scroll down and click on lower rank results.

Click **MORE** on album/track/artist and **LESS** on playlist.

**MORE likely** to save/add but **LESS likely** to stream directly.

Understanding mindset helps us understand search satisfaction.
Users ask for Spotify to play music, without saying what they would like to hear *(open mindset)*

A type of push paradigm and how it translates to the music context.

Findings from qualitative research.
Non-specific querying is a way for a user to **effortlessly** start a listening session via voice.

Non-specific querying is a way to remove the **burden of choice** when a user is open to lean-back listening.

**User education** matters as users will not engage in a use-case they do not know about.

**Trust** and control are central to a positive experience. Users need to trust the system enough to try it out.
Conclusions

Focused mindset is a typical and common case of pull paradigm. Understanding the focus mindset can inform measures of search satisfaction.

Open mindset is important for discovery and lean-back experiences.

Conversational search (Voice) allows for pull & push paradigms if done right.
Some final words
Making AI works at Spotify

Evaluation offline and online

Algorithm(s)

Training & Datasets

Interaction & feedbacks data

Features (item)

Features (user)

Features (context)

Metric(s)

Qualitative research

Business metrics
Making AI works at Spotify

... in this talk

Evaluation offline and online

ML-Lab

Algorithm(s)

BaRT

2-side marketplace

Metric(s)

Reward function for BaRT

Features (item)

Features (user)

Features (context)

Conversational search (Voice)

Focused mindset in search

Interaction & feedbacks data

Training & Datasets
Thank you!

Join the band