Two-stage Model for Automatic Playlist Continuation at Scale

Lead: Yichao Lu
Facilitators: Omar Nada, Himanshu Rai

Toronto Deep Learning Series
March 11, 2019

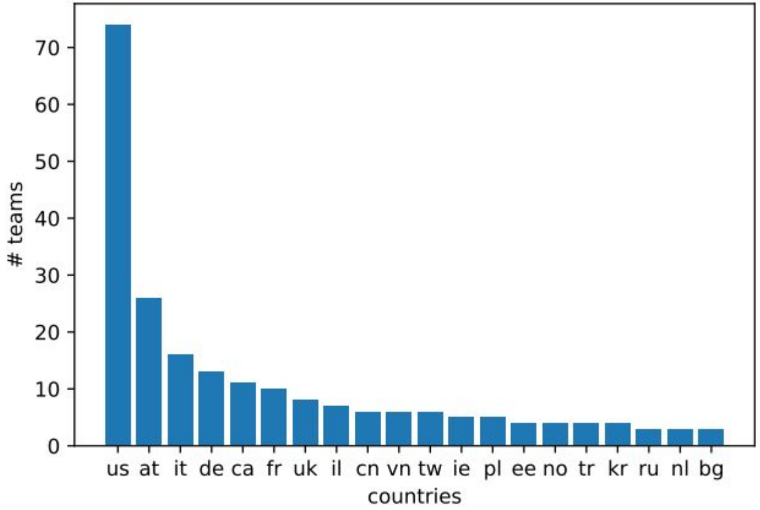
layer6

ACM RecSys Challenges

- One of the biggest annual competitions in the domain of recommender systems.
- Fervent participation from both industry as well as academia.
- Past winners include Alibaba (2016) and Yandex (2015)
- Organized in conjunction with the ACM Conference Series on Recommender Systems, where we presented our paper.
- Layer 6 has been the consecutive winner for years 2017 and 2018.

layer6

Competition Stats for 2018



- 1,791 people - 1,430 from academia and 361 from industry.
- 410 teams
- 1,467 submissions.

Source: An Analysis of Approaches Taken in the ACM RecSys Challenge 2018 for Automatic Music Playlist Continuation (https://arxiv.org/abs/1810.01520)

layer6

Spotify RecSys’18 Challenge

- Given a collection of 1M Spotify playlists with over 2M songs, create a playlist continuation model (task of sequential recommendation).

- Given a playlist of arbitrary length with some additional metadata, the task was to recommend up to 500 tracks that fit the target characteristics of the original playlist.

- This model has to work well during all stages of playlist creation – from cold start to first few songs to established playlists with many songs

- Models are evaluated on a subset of 10K playlists that have different number of songs - between 0 and 100.

layer6

Unique Aspects of Spotify RecSys Dataset

- A lot of tasks (in NLP, recommender systems, etc.) have a sequential nature and so do the datasets.
- The Million Playlist Dataset (MPD) from RecSys this year - one of the largest sequential recommender datasets.
- The learnings from this task are easily transferable to different domains e.g., language (language modelling) and vision (videos).
- At the bank we handle lots of textual and transactional data that is sequential.

layer6

Million Playlist Dataset Stats

Property	Value
Number of playlists	1,000,000
Number of tracks	66,346,428
Number of unique tracks	2,262,292
Number of unique albums	734,684
Number of unique artists	295,860
Number of unique playlist titles	92,944
Number of unique normalized playlist titles	17,381
Average playlist length (tracks)	66.35

layer6

Million Playlist Dataset


```
"playlists": [
    {
        "name": "disney",
        "collaborative": "false",
        "pid": 1000,
        "modified_at": 1457827200,
        "num_tracks": 189,
        "num_albums": 16,
        "num_followers": 1,
        "tracks": [
            {
                "pos": 0,
                "artist_name": "Original Broadway Cast - The Little Mermaid",
                "track_uri": "spotify:track:5IbCV9Icebx8rR6wAp5hhP",
                "artist_uri": "spotify:artist:3TymzPhJTMyupk7P5xkahM",
                "track_name": "Fathoms Below - Broadway Cast Recording",
                "album_uri": "spotify:album:3ULJeOMgroG27dpn27MDfS",
                "duration_ms": 154506,
                "album_name": "The Little Mermaid: Original Broadway Cast Recording"
            },
            {
                "pos": 1,
                "artist_name": "Original Broadway Cast - The Little Mermaid",
                "track_uri": "spotify:track:6rKVAvjHcxAzZ1BHtwh5yC",
                "artist_uri": "spotify:artist:3TymzPhJTMyupk7P5xkahM",
                "track_name": "Daughters Of Triton - Broadway Cast Recording",
                "album_uri": "spotify:album:3ULJeOMgroG27dpn27MDfS",
                "duration_ms": 79066,
                "album_name": "The Little Mermaid: Original Broadway Cast Recording"
            },
            {
                "pos": 2,
                "artist_name": "Original Broadway Cast - The Little Mermaid",
                "track_uri": "spotify:track:6Jlkb1Wh08RYHstWScsTvg",
                "artist_uri": "spotify:artist:3TymzPhJTMyupk7P5xkahM",
                "track_name": "The World Above - Broadway Cast Recording",
                "album_uri": "spotify:album:3ULJeOMgroG27dpn27MDfS",
                "duration_ms": 94600,
                "album_name": "The Little Mermaid: Original Broadway Cast Recording"
            },
```


layer6

Test Playlist categories

The 10,000 playlists are made up of 10 different challenge categories, with 1,000 playlists in each category:

1. Predict tracks for a playlist given its title only
2. Predict tracks for a playlist given its title and the first track
3. Predict tracks for a playlist given its title and the first 5 tracks
4. Predict tracks for a playlist given its first 5 tracks (no title)
5. Predict tracks for a playlist given its title and the first 10 tracks
6. Predict tracks for a playlist given its first ten tracks (no title)
7. Predict tracks for a playlist given its title and the first 25 tracks
8. Predict tracks for a playlist given its title and 25 random tracks
9. Predict tracks for a playlist given its title and the first 100 tracks
10. Predict tracks for a playlist given its title and 100 random tracks

layer6

Our Approach

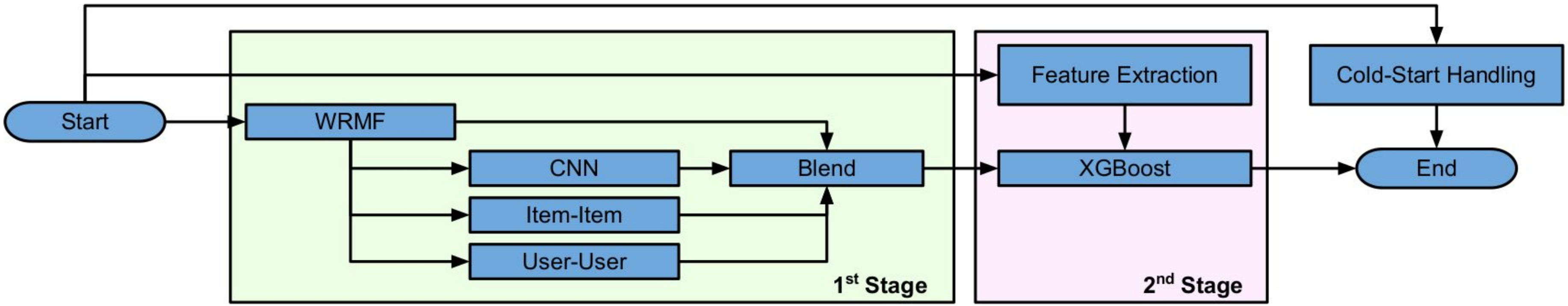
- Two stage architecture:

    - First stage reduces large 2.2M song search space to a small set of candidates with high recall.
    - Second stage re-ranks the candidates maximizing precision at the top.
    - Cold start playlists are handled separately with a cold start model.

- We use an ensemble of several models in the first stage and pairwise gradient boosting model in the second stage.

layer6

Architecture Diagram



- To encourage simplicity and improve generalization the same pipeline is applied to all test playlists.

layer6

Latent factor models(First Stage)

- Latents encode useful information about users and items and provide efficient inference and high accuracy.

- We use WRMF[Hu et al., 2008] as our base method which incorporates implicit feedback through confidences:

$$arg \min_{U,V} \sum_{i,j} c_{ij}(R_{i,j} - U_i V_j)^2 + \lambda_U \|U_i\|_2^2 + \lambda_V \|V_j\|_2^2$$

where confidence $c_{ij} = 1 + \alpha R_{i,j}$ and $R_{i,j}$ are the ratings in the playlist-song matrix.

- WRMF is used to retrieve a subset of 20K songs for each playlists, other models' scores are only computed on this subset. Gives over 90% recall at 20K.

- We use a deep CNN model to encode sequential information, which is missing in other CF methods like user-user and item-item as well as wrmf.

layer6

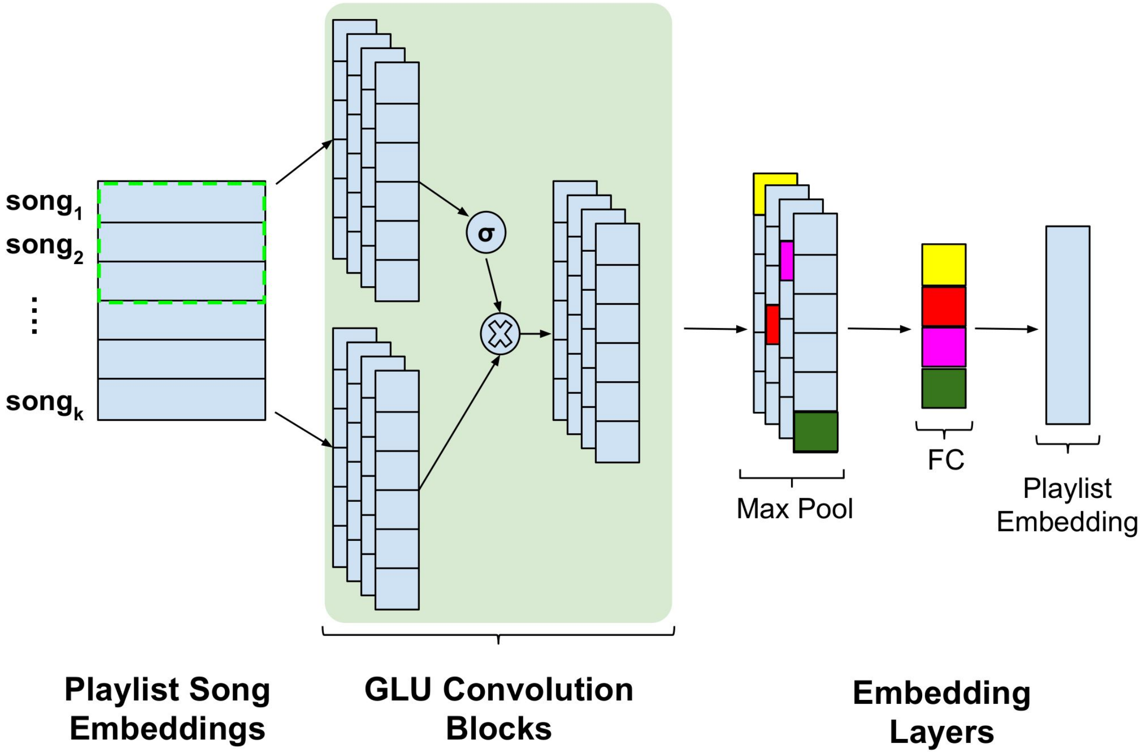
CNN Embedding Model

- Most CF models ignore temporal order in user-item interactions
- Temporal order often contains highly useful information for personalization
- Recent work in temporal CF has largely focused on RNN models
- However, RNNs are inherently sequential and are thus slow during both training and inference

CNN Embedding Model

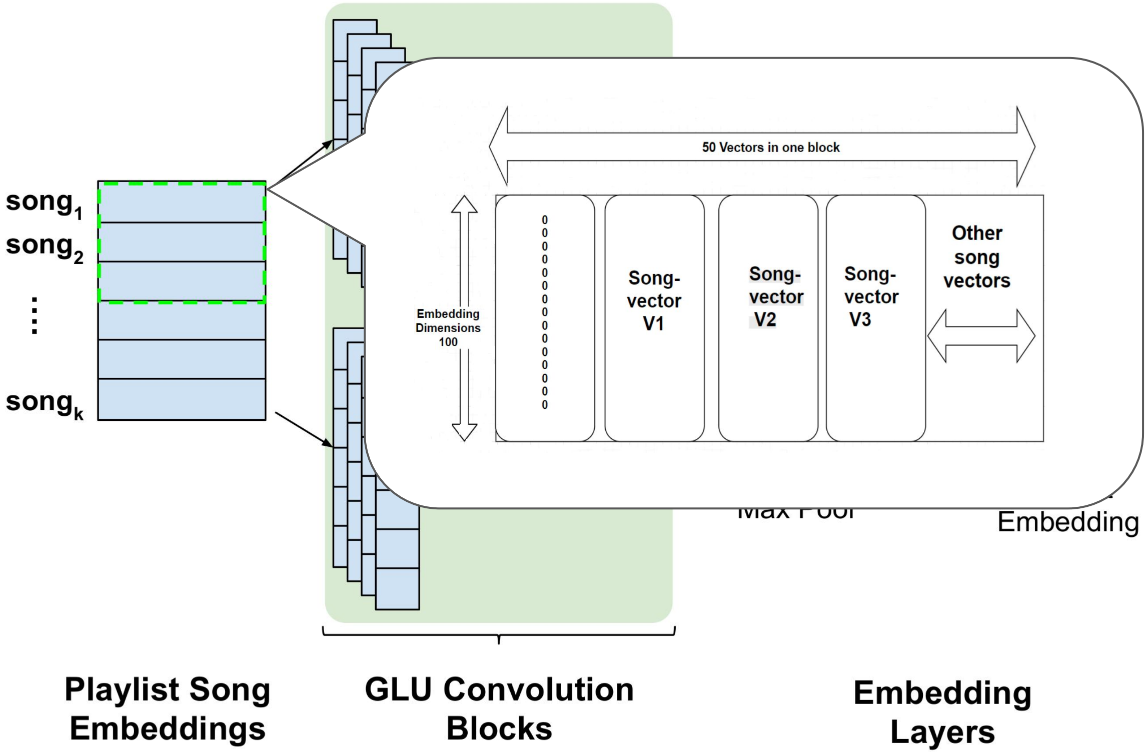
- We have been focusing on applying CNN models to temporal CF data
- CF is similar to language modeling: **playlist=document** and **song=word**
- CNN models have recently been shown to perform extremely well on language modeling tasks [Dauphin at al., 2017]
- We extend ideas from language modeling and develop a CNN model for temporal CF

CNN Embedding Model



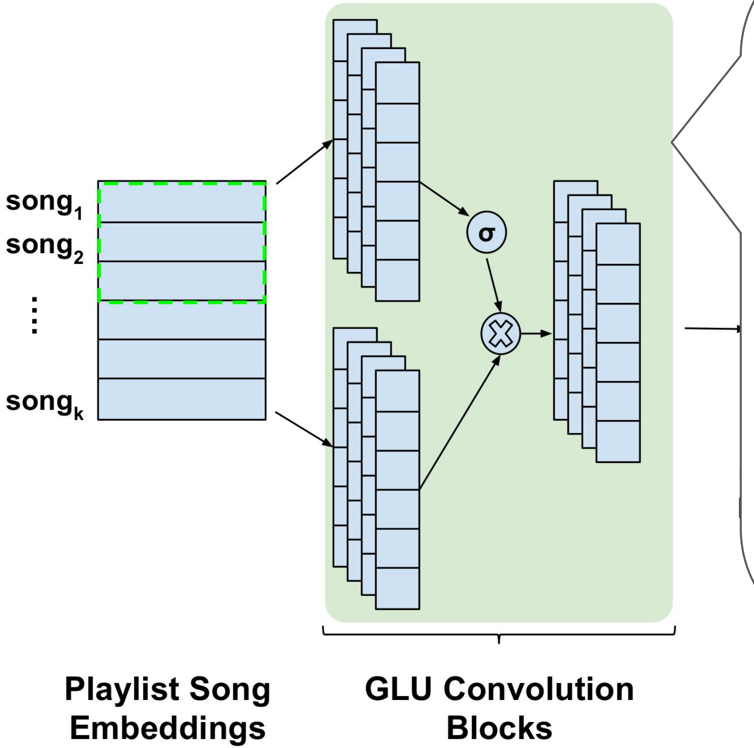
layer6

CNN Embedding Model



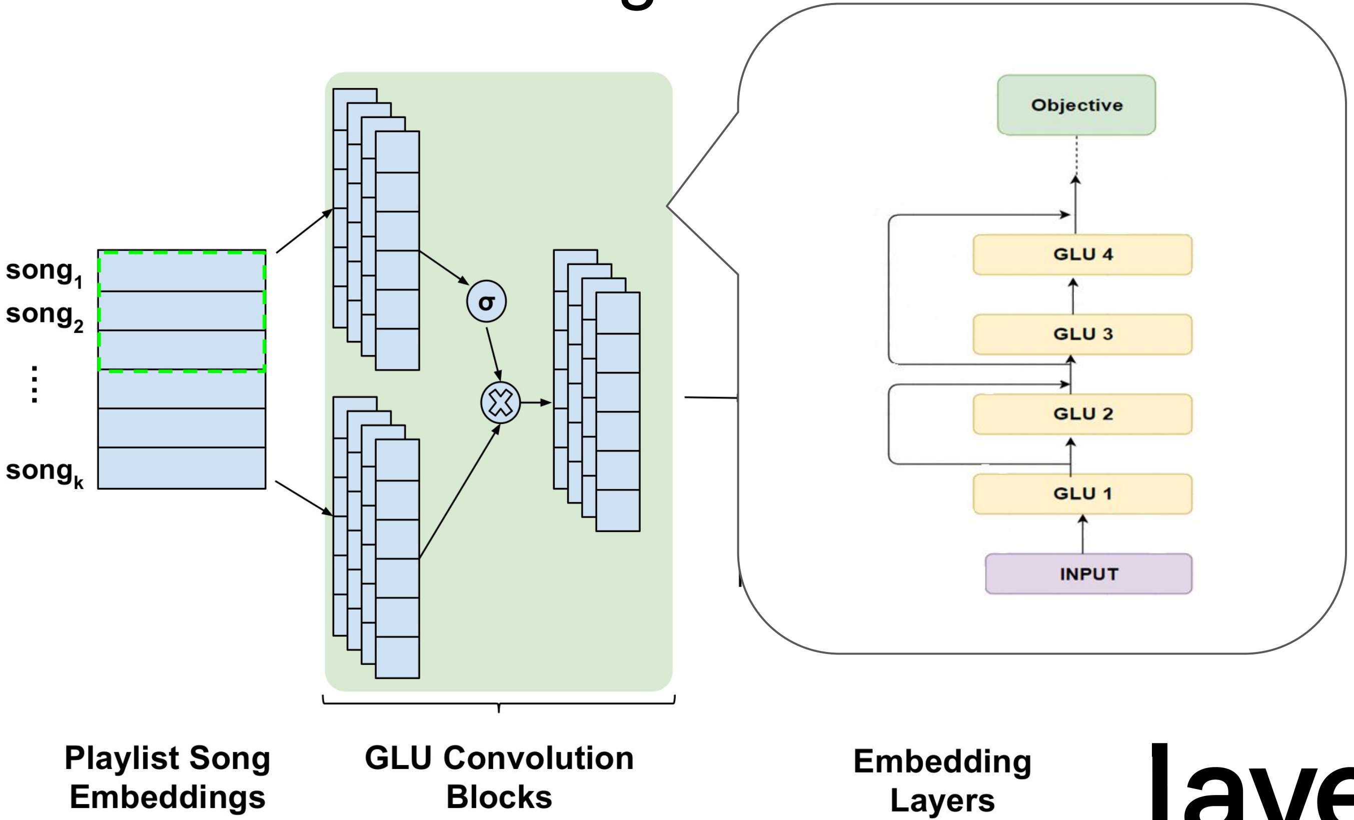
layer6

CNN Embedding Model



- GLUs(Dauphin et. al.) solve the problem of vanishing gradients in deep nets.
- This gating is analogous to gating in an lstm.
- $h^l(x) = (h^{l-1}(x) * W^l + b^l) \otimes \sigma(h^{l-1}(x) * V^l + c^l)$
- Added dropout layer for regularization.
- Adding batch norm further improved the results.

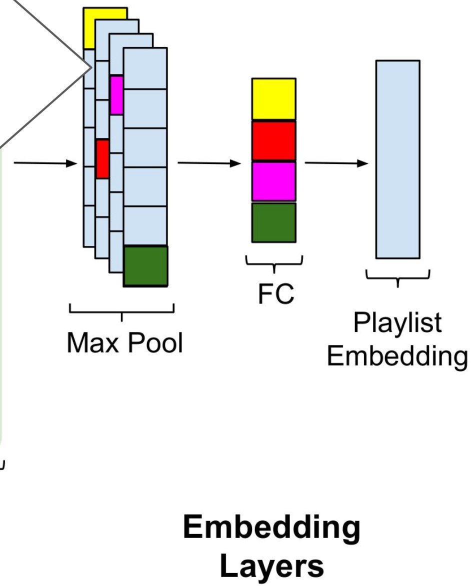
CNN Embedding Model



layer6

CNN Embedding Model

- Use of max pool/top-k pool along channels
- This makes up for variable input lengths.
- The best result involves 900 kernels for each convolution and 7 GLU blocks.
- We update the initial latents on each epoch.
- Used adam optimizer with plateau decay learning scheme.



layer6

Training Objective Formulation

- Playlist formed by concatenation of song vectors : $\mathbf{\Phi_i^{1:k}} = [\mathbf{V_1^{cnn}}, \mathbf{V_2^{cnn}}, \ldots, \mathbf{V_k^{cnn}}]$
- Playlist embedding obtained after passing through CNN: $\mathbf{U_i^{cnn}} = f(\mathbf{\Phi_i^{1:k}}, \theta)$
- Probability that a song j is present in a playlist i is, $P(\mathbf{V_j^{cnn}} | \mathbf{U_i^{cnn}}) = \frac{\mathbf{1}}{\mathbf{1+e^{-U_i^{cnn} V_j^{cnn}}}}$
- Raising the probability of songs that follow $\mathbf{V_k^{cnn}}$ and lowering for all other songs, the loss :

$$-\sum_{\substack{j \in \mathcal{V}(i) \\ j>k}} \log\left(\frac{1}{1+e^{-\mathbf{U_i^{cnn} V_j^{cnn}}}}\right) - \sum_{j' \notin \mathcal{V}(i)} \log\left(1 - \frac{1}{1+e^{-\mathbf{U_i^{cnn} V_{j'}^{cnn}}}}\right)$$

- Predicting next 3 songs gave better results than just 1 song. We keep the number of negative samples to be 10.

layer6

CNN Embedding Model

- Conceptually “good” playlist embedding is an accurate predictor of songs that follow

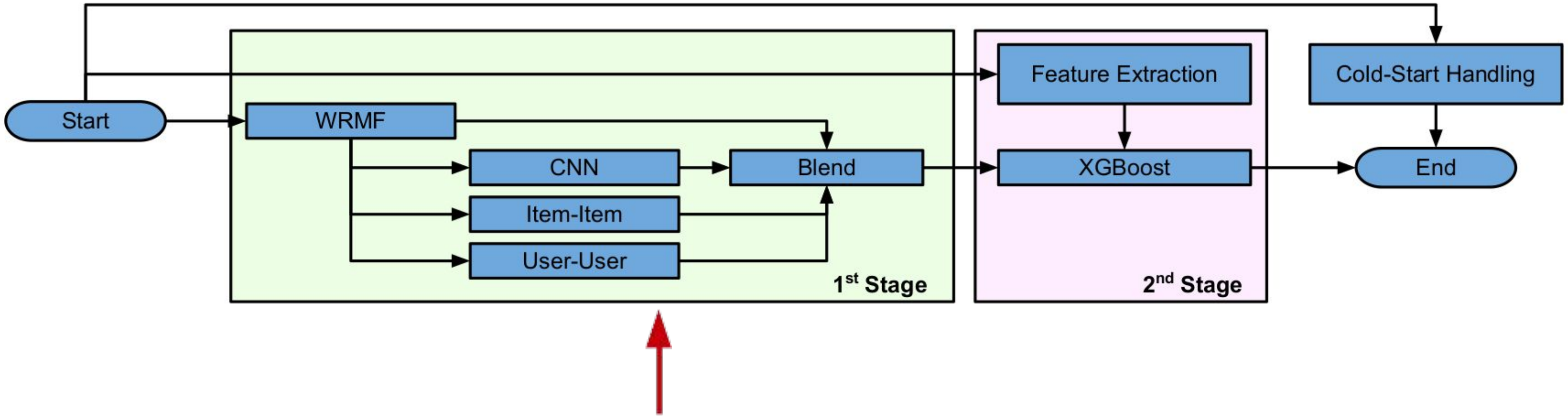
- During training we repeat these steps:
    - Sample subsequence of songs from a playlist
    - Embed this subsequence with CNN
    - Use embedding to predict next songs

- Training time is 2x-5x faster than RNNs

- CNN learns to generate accurate order dependent embeddings that can be used for fast retrieval

layer6

First Stage Summary



- Scores from the four first stage models are blended together and passed to the second stage

Second Stage Model

- Only applied to a small subset of songs retrieved by first stage so prioritize accuracy over efficiency
- Use gradient boosting to directly map playlist-song pairs to relevance scores
- The goal is to learn pairwise correlations that are difficult to capture in the first stage

layer6

Second Stage Model

- Most effort in this stage was spent on feature engineering

- Four feature groups that capture both interaction and content information:

  - **Input From First Stage**: scores from each model + blend
  - **Playlist Features**: name, length, song/artist/album stats, within playlist song similarity etc.
  - **Song Features**: duration, artist/album stats, playlist stats where this song appears etc.
  - **Playlist-Song features**: sim. between song and songs in playlist, analogous sim. for playlists

Second Stage Model

- Given 20K candidate songs produced by the first stage we train second stage to push relevant songs to the top

- Randomly sample (up to) 20 positive and 20 “hard” negative songs for each playlist

- Train GBM model with pairwise ranking objective.

layer6

Cold Start Model

- Used for the 1000 playlists in the test set with just playlist names.
- We form R^{name} matrix-Songs and playlists concatenated together as rows. Columns corresponding to names.
- Factorize using truncated SVD to obtain U^{name} and V^{name} matrices. Used these vectors for retrieval.

layer6

2 minutes break……

Results

- Custom Java stack with data structures designed specifically for CF.
- Training both stages takes approximately 12 hours to achieve leaderboard accuracy.
- End-to-end inference can be done in under 50ms per playlist.

layer6

Results

- Validation set results

	RPREC	NDCG	Clicks
WRMF	0.1641	0.3350	2.1230
CNN	0.1594	0.3230	2.1157
Item-Item	0.1772	0.3534	2.1649
User-User	0.1768	0.3550	1.6332
Blend	0.1866	0.3728	1.4064
2^{nd} Stage	**0.1985**	**0.3846**	**1.1998**
Cold Start			
Popular	0.0395	0.0815	17.2662
Name SVD	0.1106	0.2083	7.9609

layer6

Results

- Validation set results

	RPREC	NDCG	Clicks
WRMF	0.1641	0.3350	2.1230
CNN	0.1594	0.3230	2.1157
Item-Item	0.1772	0.3534	2.1649
User-User	0.1768	0.3550	1.6332
Blend	0.1866	0.3728	1.4064
2^{nd} Stage	**0.1985**	**0.3846**	**1.1998**
Cold Start			
Popular	0.0395	0.0815	17.2662
Name SVD	0.1106	0.2083	7.9609

- Blend significantly improves performance of the first stage

$$S^{blend} = w_1 S^{wrmf} + w_2 S^{CNN} + w_3 S^{user} + w_4 S^{item}$$

(w1 = 0.1, w2 = 0.4, w3 =0.3, w4 = 0.3)

layer6

Results

- Validation set results

	RPREC	NDCG	Clicks
WRMF	0.1641	0.3350	2.1230
CNN	0.1594	0.3230	2.1157
Item-Item	0.1772	0.3534	2.1649
User-User	0.1768	0.3550	1.6332
Blend	0.1866	0.3728	1.4064
2^{nd} Stage	**0.1985**	**0.3846**	**1.1998**
Cold Start			
Popular	0.0395	0.0815	17.2662
Name SVD	0.1106	0.2083	7.9609

- CNN contributes the most to the blend so order is very important!

layer6

Results

- Validation set results

	RPREC	NDCG	Clicks
WRMF	0.1641	0.3350	2.1230
CNN	0.1594	0.3230	2.1157
Item-Item	0.1772	0.3534	2.1649
User-User	0.1768	0.3550	1.6332
Blend	0.1866	0.3728	1.4064
2^{nd} Stage	**0.1985**	**0.3846**	**1.1998**
Cold Start			
Popular	0.0395	0.0815	17.2662
Name SVD	0.1106	0.2083	7.9609

- Second stage consistently adds over a point in both RPREC and NDCG

layer6

Results

- Validation set results

	RPREC	NDCG	Clicks
WRMF	0.1641	0.3350	2.1230
CNN	0.1594	0.3230	2.1157
Item-Item	0.1772	0.3534	2.1649
User-User	0.1768	0.3550	1.6332
Blend	0.1866	0.3728	1.4064
2^{nd} Stage	**0.1985**	**0.3846**	**1.1998**
Cold Start			
Popular	0.0395	0.0815	17.2662
Name SVD	0.1106	0.2083	7.9609

- We compare our name-based SVD approach to a simple popularity baseline.
- To conduct this evaluation we remove all training songs from the validation playlists and treat them as cold start.

Results

- Leaderboard results

	RPREC	NDCG	Clicks	Borda
vl6	**0.2241**	**0.3946**	1.7839	**329**
hello world!	0.2233	0.3932	1.8952	323
Avito	0.2153	0.3845	**1.7818**	322
Creamy Fireflies	0.2201	0.3856	1.9335	320
MIPT_MSU	0.2167	0.3823	1.8754	320

Results

- Leaderboard results

	RPREC	NDCG	Clicks	Borda
vl6	**0.2241**	**0.3946**	1.7839	**329**
hello world!	0.2233	0.3932	1.8952	323
Avito	0.2153	0.3845	**1.7818**	322
Creamy Fireflies	0.2201	0.3856	1.9335	320
MIPT_MSU	0.2167	0.3823	1.8754	320

  - Outperform all teams on RPREC and NDCG, and second best on Clicks

Results

- Qualitative Results

Playlist 1	Playlist 2	Playlist 3
1. Drake - Sneakin	1. Stone & Van Linden - Summerbreeze	1. London Philharmonic Orchestra - Symphony No. 40 in G Minor, K. 550: Allegro molto
2. Logic - The Incredible True Story	2. Havana Brown - We Run The Night	2. London Philharmonic Orchestra - Requiem, K. 626: Lacrimosa dies illa
3. Travis Scott - Birds In The Trap Sing McKnight	3. Calvin Harris - Sweet Nothing	3. Elisabeth Ganter - Concerto in A Major for Clarinet and Orchestra, K. 622: II. Adagio
4. Big Sean - I Decided	4. David Guetta - Titanium	4. Tbilisi Symphony Orchestra - Symphony No.25 In G Minor, K. 183 I. Allegro Con Brio
5. The Weeknd - Starboy	5. Calvin Harris - Feel So Close	5. Tbilisi Symphony Orchestra - Requiem Mass In D Minor, K. 626 : II. Dies Irae
	Model Recommendations	
Drake - Fake Love	Rihanna - We Found Love	London Philharmonic Orchestra - String Quintet No.4 in G Minor, K.516: I. Allegro
Migos - Bad and Boujee	Swedish House Mafia - Don't You Worry Child	Takács Quartet- String Quintet No.3 in C Minor, K.515: III. Andante
Post Malone - Congratulations	Calvin Harris - Summer	Columbia Symphony Orchestra - Serenade in G major, K. 525 i. Allegro

- Model effectively preserves playlist themes (EDM, Hip Hop, Classical) and recommends diverse artists

layer6

Results

- Qualitative Results

Playlist 1	
1. Drake - Sneakin	G Minor, K. 550: Allegro molto
2. Logic - The Incredible True Story	crimosa dies illa
3. Travis Scott - Birds In The Trap Sing Mc	d Orchestra, K. 622: II. Adagio
4. Big Sean - I Decided	inor, K. 183 I. Allegro Con Brio
5. The Weeknd - Starboy	or, K. 626 : II. Dies Irae
Drake - Fake Love	G Minor, K.516: I. Allegro
Migos - Bad and Boujee	I. Andante
Post Malone - Congratulations	. 525 i. Allegro

Playlist 1
1. Drake - Sneakin
2. Logic - The Incredible True Story
3. Travis Scott - Birds In The Trap Sing McKnight
4. Big Sean - I Decided
5. The Weeknd - Starboy

Drake - Fake Love
Migos - Bad and Boujee
Post Malone - Congratulations

Results

- Qualitative Results

Playlist 1	Playlist 2
1. Drake - Sneakin	1. Stone & Van Linden - Su
2. Logic - The Incredible True Story	2. Havana Brown - We Run The Nig
3. Travis Scott - Birds In The Trap Sing McKnight	3. Calvin Harris - Sweet Nothing
4. Big Sean - I Decided	4. David Guetta - Titanium
5. The Weeknd - Starboy	5. Calvin Harris - Feel So Close
	Model R
Drake - Fake Love	Rihanna - We Found Love
Migos - Bad and Boujee	Swedish House Mafia - Don't You Wor
Post Malone - Congratulations	Calvin Harris - Summer

Playlist 2
1. Stone & Van Linden - Summerbreeze
2. Havana Brown - We Run The Night
3. Calvin Harris - Sweet Nothing
4. David Guetta - Titanium
5. Calvin Harris - Feel So Close

Rihanna - We Found Love
Swedish House Mafia - Don't You Worry Child
Calvin Harris - Summer

layer6

Results

- Qualitative Results

Playlist 1 Playlist 2 Playlist 3

Playlist 3

1. London Philharmonic Orchestra - Symphony No. 40 in G Minor, K. 550: Allegro molto
2. London Philharmonic Orchestra - Requiem, K. 626: Lacrimosa dies illa
3. Elisabeth Ganter - Concerto in A Major for Clarinet and Orchestra, K. 622: II. Adagio
4. Tbilisi Symphony Orchestra - Symphony No.25 In G Minor, K. 183 I. Allegro Con Brio
5. Tbilisi Symphony Orchestra - Requiem Mass In D Minor, K. 626 : II. Dies Irae

London Philharmonic Orchestra - String Quintet No.4 in G Minor, K.516: I. Allegro
Takács Quartet- String Quintet No.3 in C Minor, K.515: III. Andante
Columbia Symphony Orchestra - Serenade in G major, K. 525 i. Allegro

nic Orchestra - Symphony No. 40 in G Minor, K. 550: Allegro molto
nic Orchestra - Requiem, K. 626: Lacrimosa dies illa
oncerto in A Major for Clarinet and Orchestra, K. 622: II. Adagio
Symphony No.25 In G Minor, K. 183 I. Allegro Con Brio
n Mass In D Minor, K. 626 : II. Dies Irae
c Orchestra - String Quintet No.4 in G Minor, K.516: I. Allegro
g Quintet No.3 in C Minor, K.515: III. Andante
Orchestra - Serenade in G major, K. 525 i. Allegro

layer6

Discussions

- Deep Learning in Recommender Systems
- CNNs vs RNNs
- One Model To Learn Them All
- Future Works

Thank You
We're hiring

careers@layer6.ai

Paper:
Two-stage Model for Automatic Playlist Continuation at Scale

Maksims Volkovs, Himanshu Rai, Zhaoyue Cheng, Ga Wu, Yichao Lu, and Scott Sanner
RecSys-2018: ACM Conference on Recommender Systems