Mostra:
A Flexible Balancing Framework to Trade-off User, Artist and Platform Objectives for Music Sequencing


Emanuele Bugliarello
emanuele@di.ku.dk
ebugliarello


R. Mehrotra

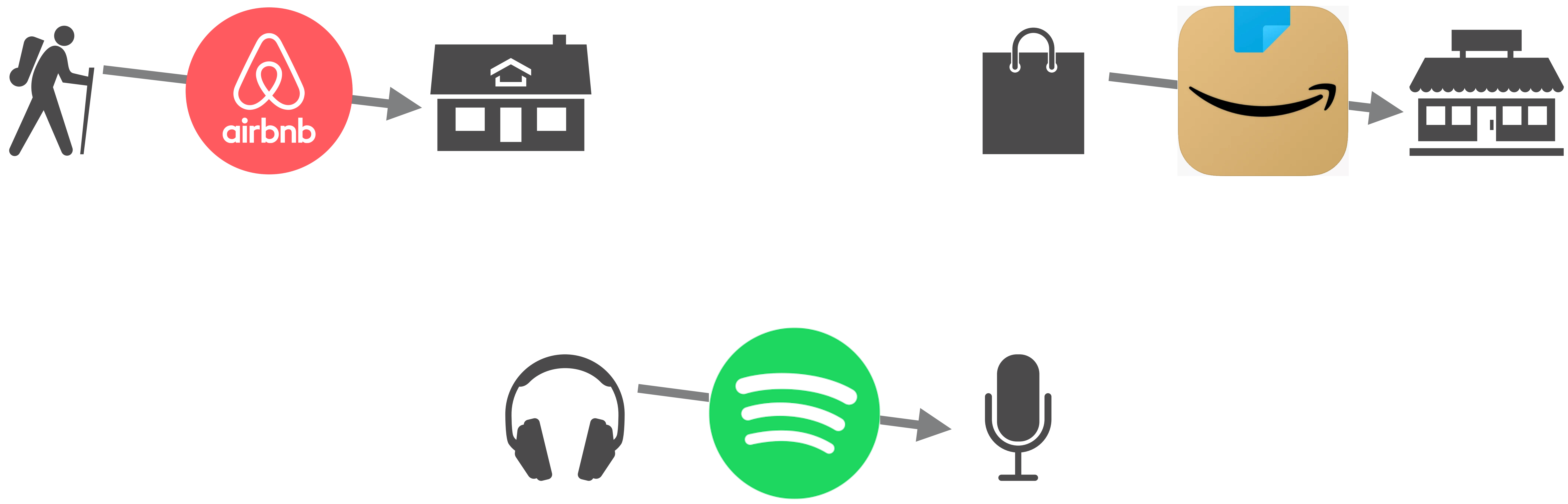

J. Kirk


M. Lalmas

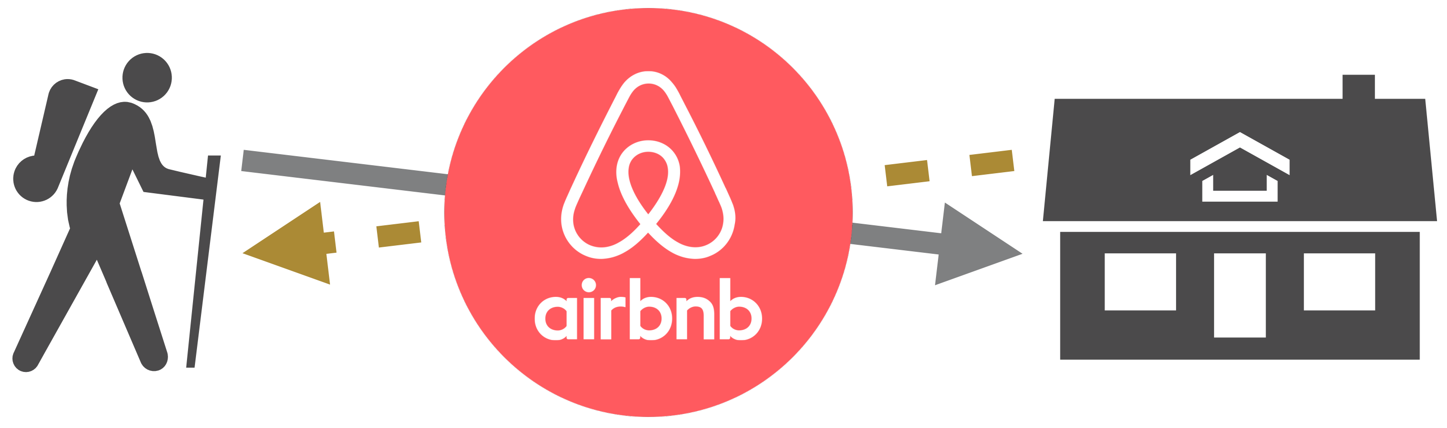
THE WEB CONFERENCE ACM
25-29 April 2022 | Lyon, France

UNIVERSITY OF COPENHAGEN
Spotify

Beyond User-centric Recommendations



Beyond User-centric Recommendations



TL; DW

🎯 **Multi-Objective Music Sequencing**
Rank songs from a large set of candidates to meet user-and creator-centric objectives

🔎 **Multiple Objectives in Music Recommendation**

Potential and difficulty of jointly satisfying multi-stakeholder goals

🤖 **Mostra**

Transformer-based encoder-decoder framework for flexible recommendations

With a novel beam search algorithm

- submodular multi-objective scoring
- counterfactual performance on user metrics

Motivation

Problem Setup

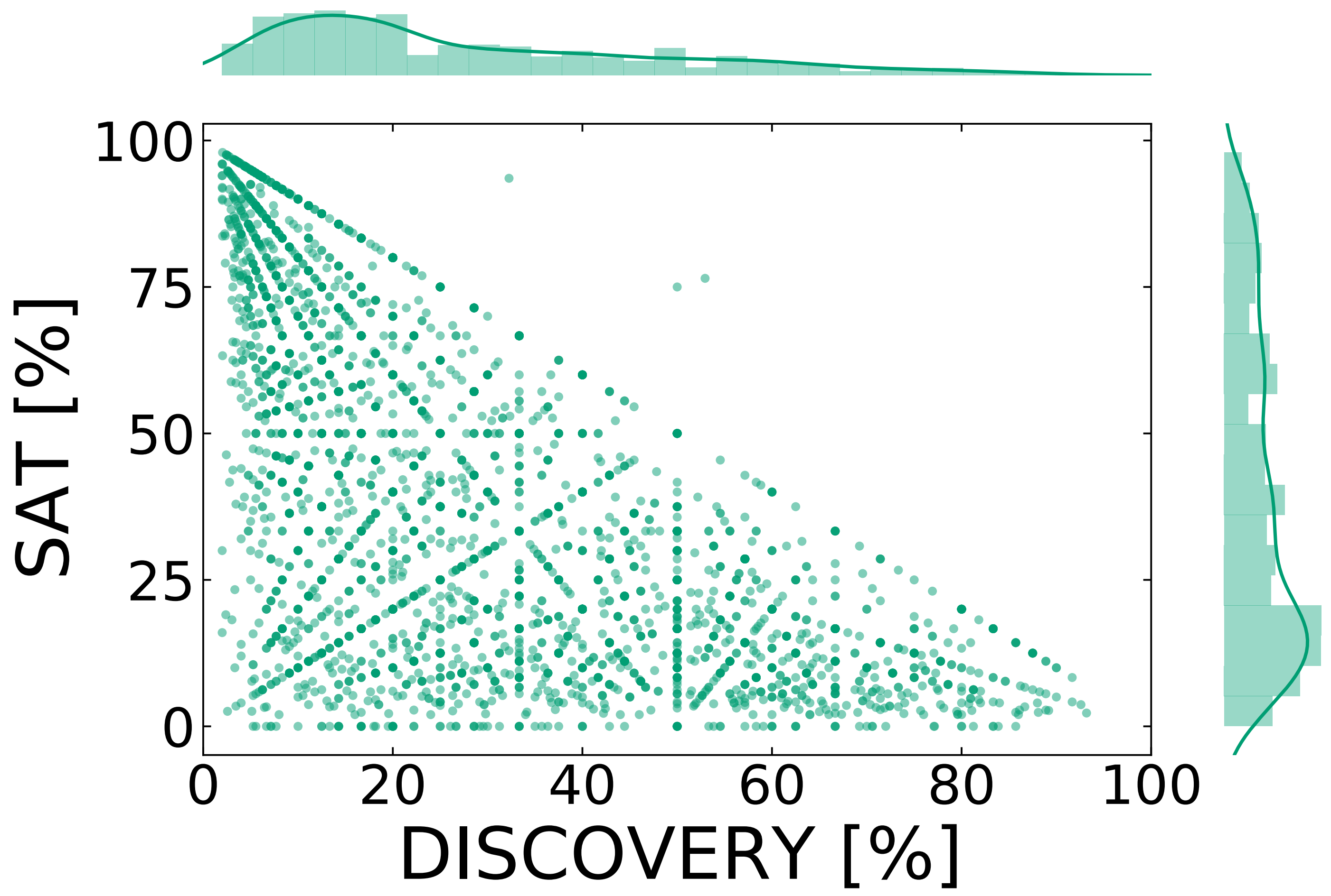
🗂️ Data Context

- Spotify radio-like music streaming sessions
- 10M users
- 500M sessions
- 1B interactions
- 7 days

🎳 (Binary) Objectives

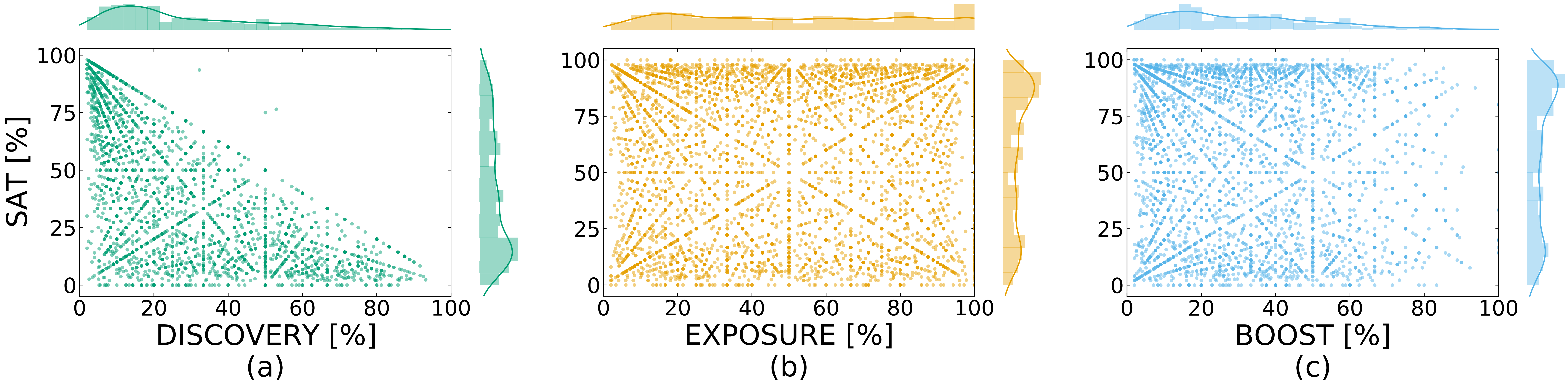
- Short-term user satisfaction (SAT): prediction of user–track completion probability
- Boost: whether a track is being boosted by the platform for strategic importance
- Exposure: whether a track belongs to an emergent artist
- Discovery: whether a track belongs to an artist a given user has never listened to

Objectives Interplay

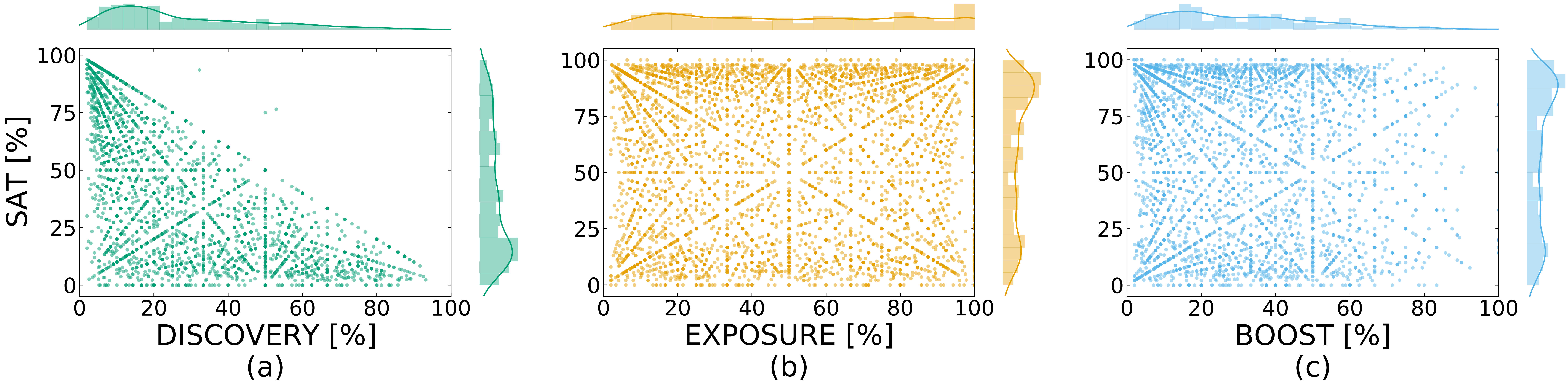


(a)

Objectives Interplay



Objectives Interplay



👀 **Objectives correlation with SAT**

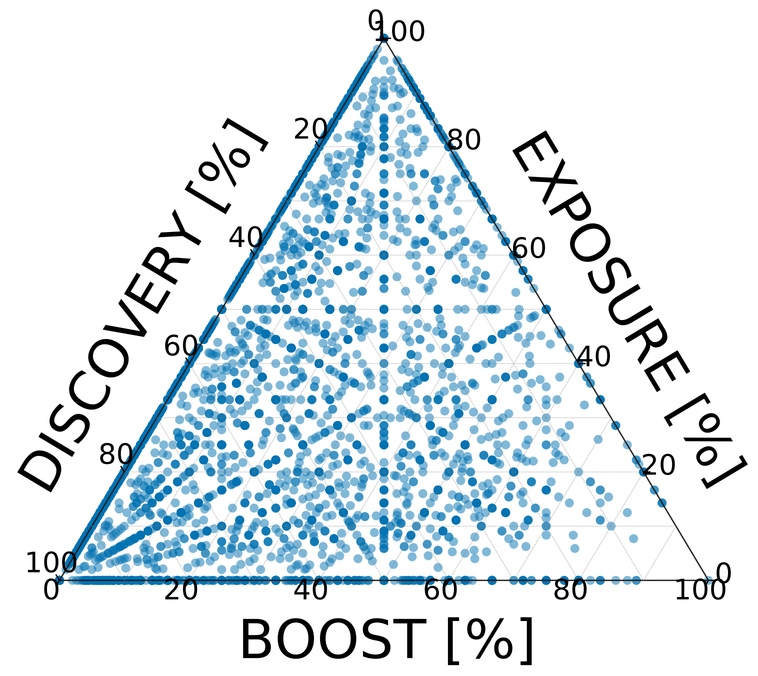
- Boost & Exposure are not correlated with user satisfaction
- Discovery has strict negative correlation with user satisfaction

Objectives Co-occurrence

📊 Plotting the % songs in a session from different objectives

😮 **Music sessions are very diverse**

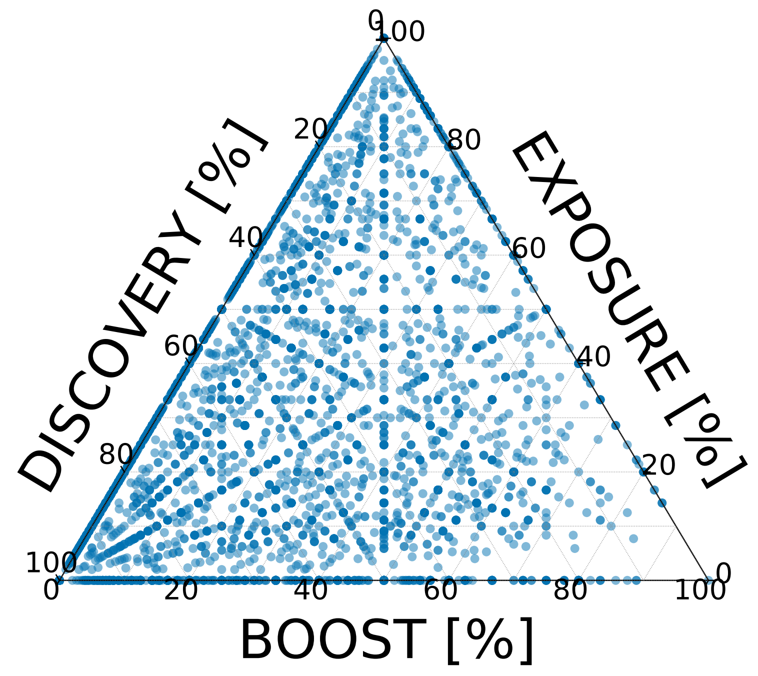
- Some sessions only have songs belonging to one objective
- Many sets have songs belonging to each of these objectives



👉 We anticipate more severe competition across objectives in certain sessions than others

Objectives Co-occurrence

📊 Plotting the % songs in a session from different objectives

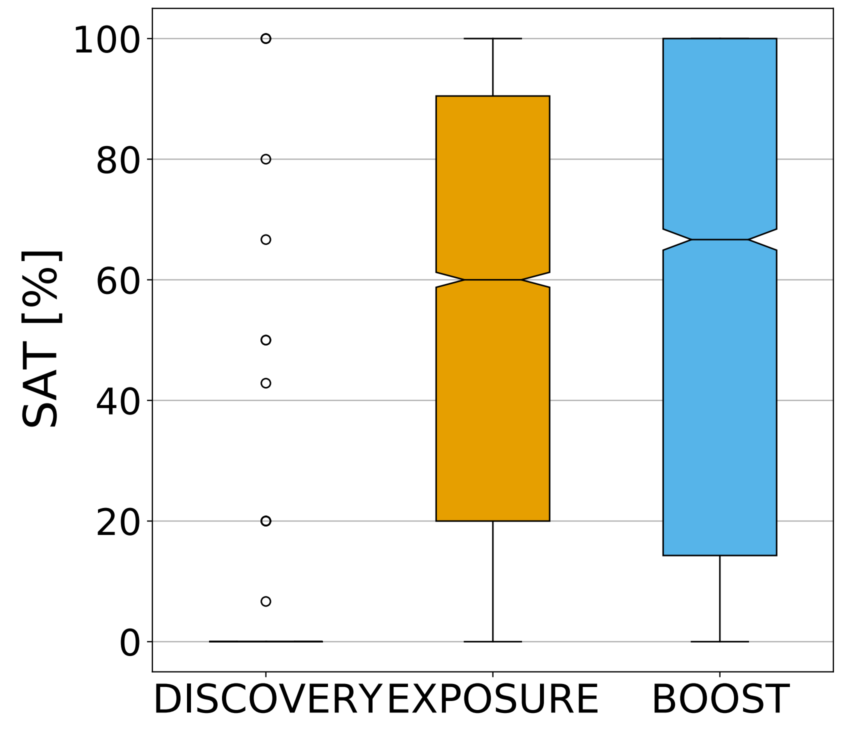


😮 **Music sessions are very diverse**

- Some sessions only have songs belonging to one objective
- Many sets have songs belonging to each of these objectives

👉 We anticipate more severe competition across objectives in certain sessions than others

📊 Plotting the SAT metric distribution for each objective



😵 **User satisfaction varies across tracks**

- SAT varies significantly across tracks with Exposure & Boost
- Discovery tracks usually have a much lower SAT

👉 Selecting the right creator-centric songs can keep SAT high

Mostra: Multi-objective Set Transformer with Counterfactual & Submodular Beam Search

Approach Overview

🎯 **Multi-Objective Music Sequencing**

Rank songs from a large set of candidates to meet user-and creator-centric objectives

🔎 **Desiderata**

- Set-aware method
- Multi-Objective decision making
- **Dynamic & flexible control** for changing business strategy needs

Approach Overview

🎯 **Multi-Objective Music Sequencing**

Rank songs from a large set of candidates to meet user-and creator-centric objectives

🔎 **Desiderata**

- Set-aware method
- Multi-Objective decision making
- **Dynamic & flexible control** for changing business strategy needs

🤖 **Mostra: M**ulti-**o**bjective **S**et **Tra**nsformer

Transformer-based encoder-decoder framework for flexible recommendations

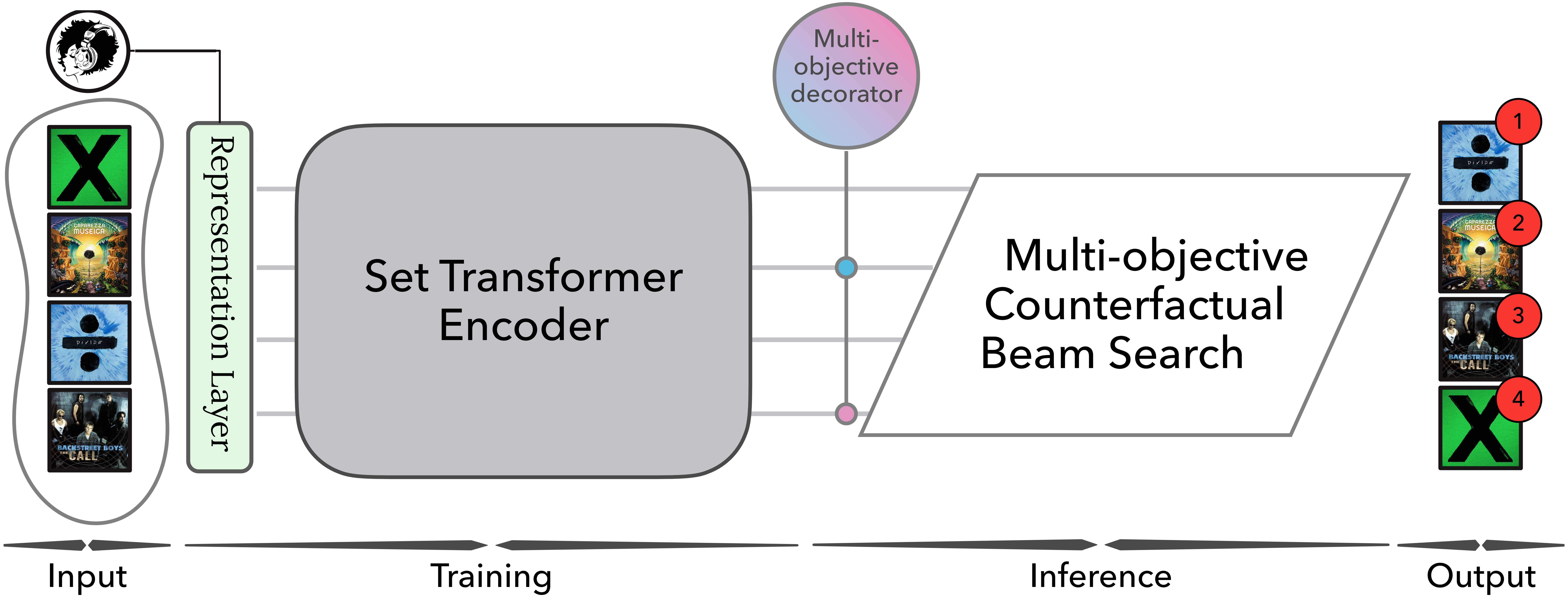
With a novel beam search algorithm

- submodular multi-objective scoring
- counterfactual performance on user metrics

Mostra: Overview

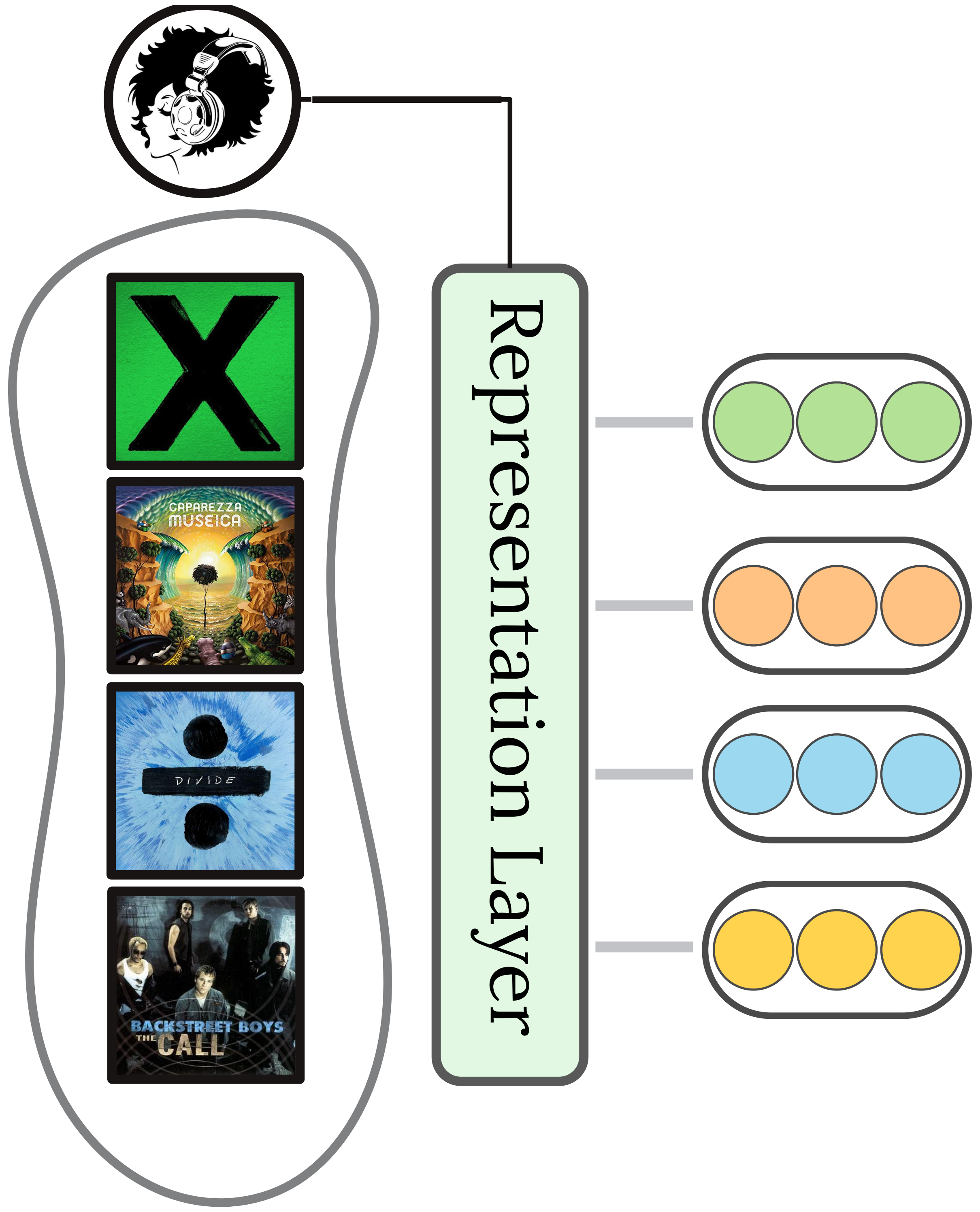
🤖 **Mostra: M**ulti-**o**bjective **S**et **Tra**nsformer

Transformer model for flexible recommendations through multi-objective counterfactual decoding



Mostra: Representation Layer

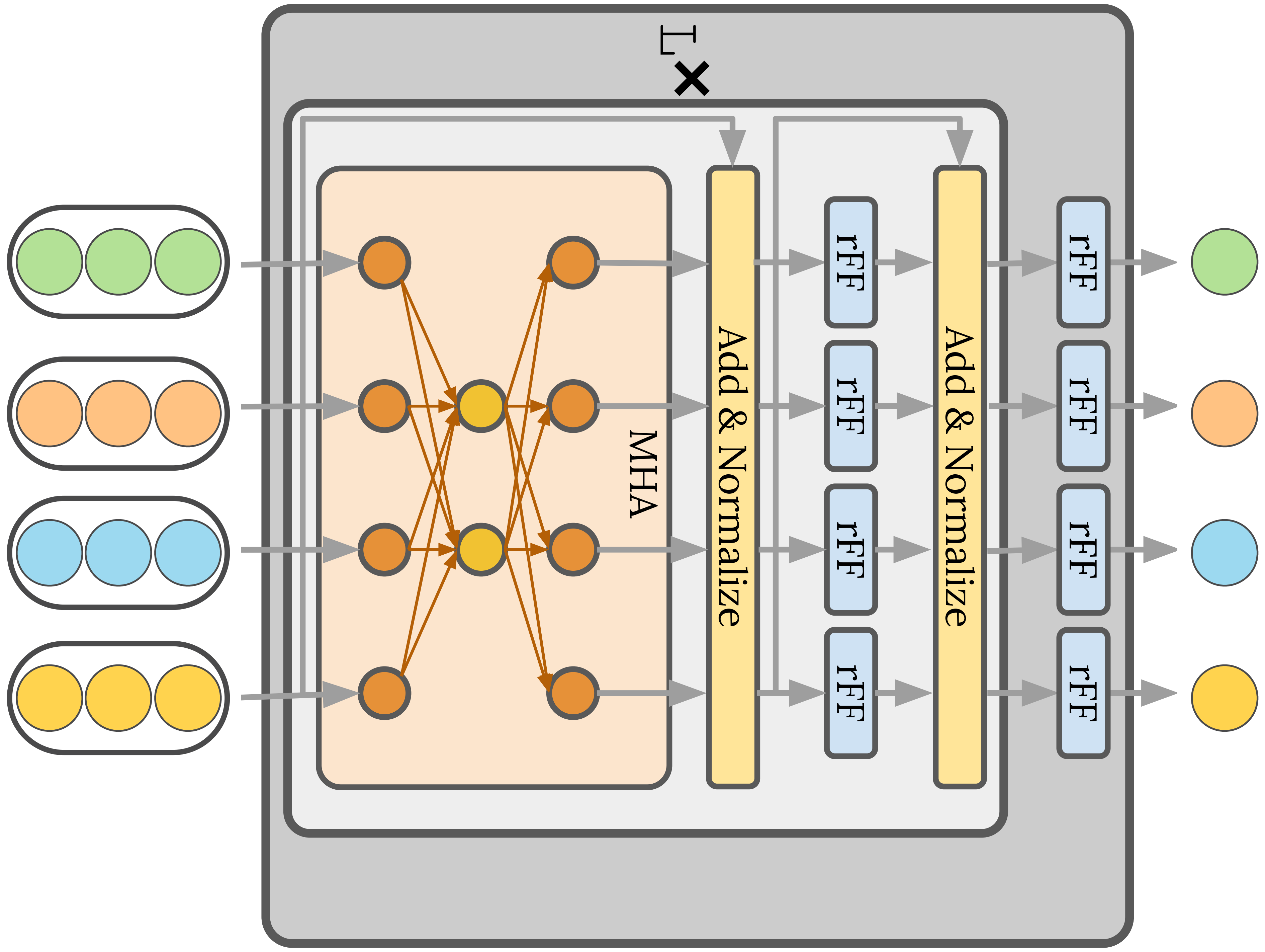
Simple module mapping User, Track and Joint features onto a vector space



Type	Feature	Description
User	embedding	40 dimensional learnt word2vec vector of user
	country	country of registration for user
Track	embedding	embedding & 40 dimensional learnt word2vec vector of track
	popularity	normalised popularity of the track
	genres	genres relevant to the track
	acoustic	16 derived acoustic features
	track length	track duration in seconds
User–Track	similarity	user–track embeddings cosine similarity
	distance	user–track embeddings Euclidean distance
	genre affinity	affinity for highest user–track overlapping genre
Session	session ID	unique session identifier for learning embeddings

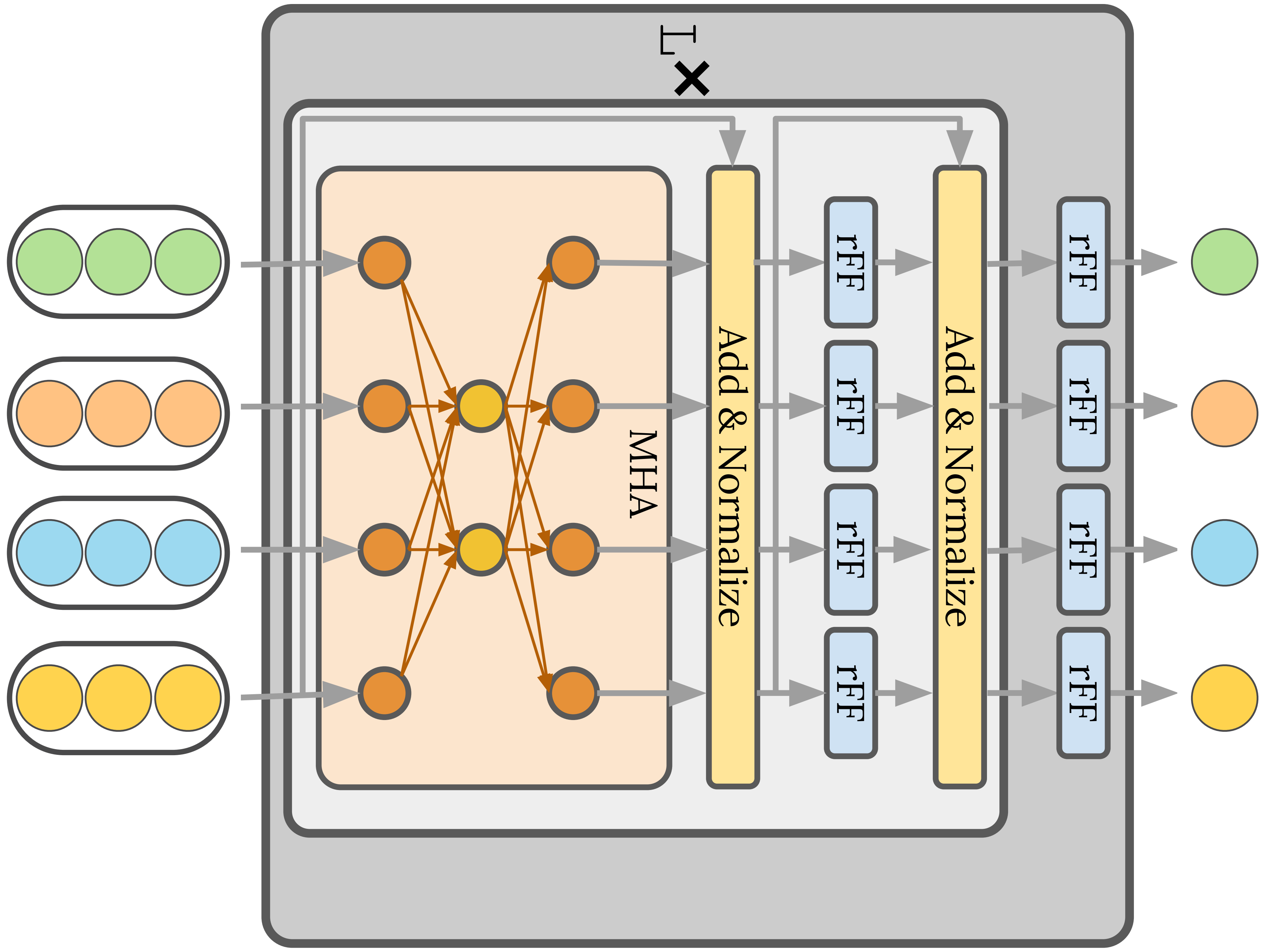
Mostra: Encoder

Set Transformer encoder giving contextualised features for each track in the pool through L layers



Mostra: Encoder

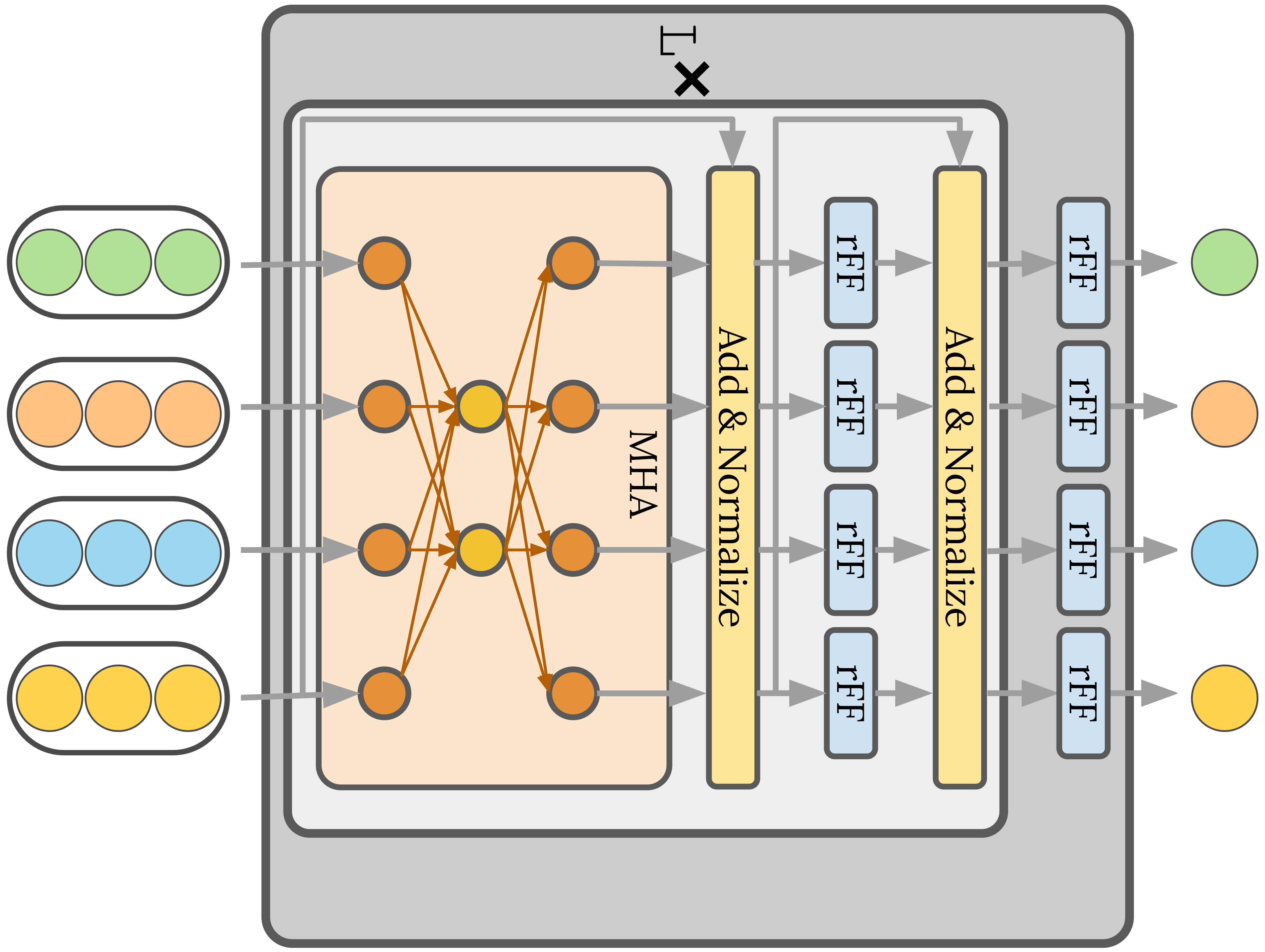
Set Transformer encoder giving contextualised features for each track in the pool through L layers



Steps:

Mostra: Encoder

Set Transformer encoder giving contextualised features for each track in the pool through L layers

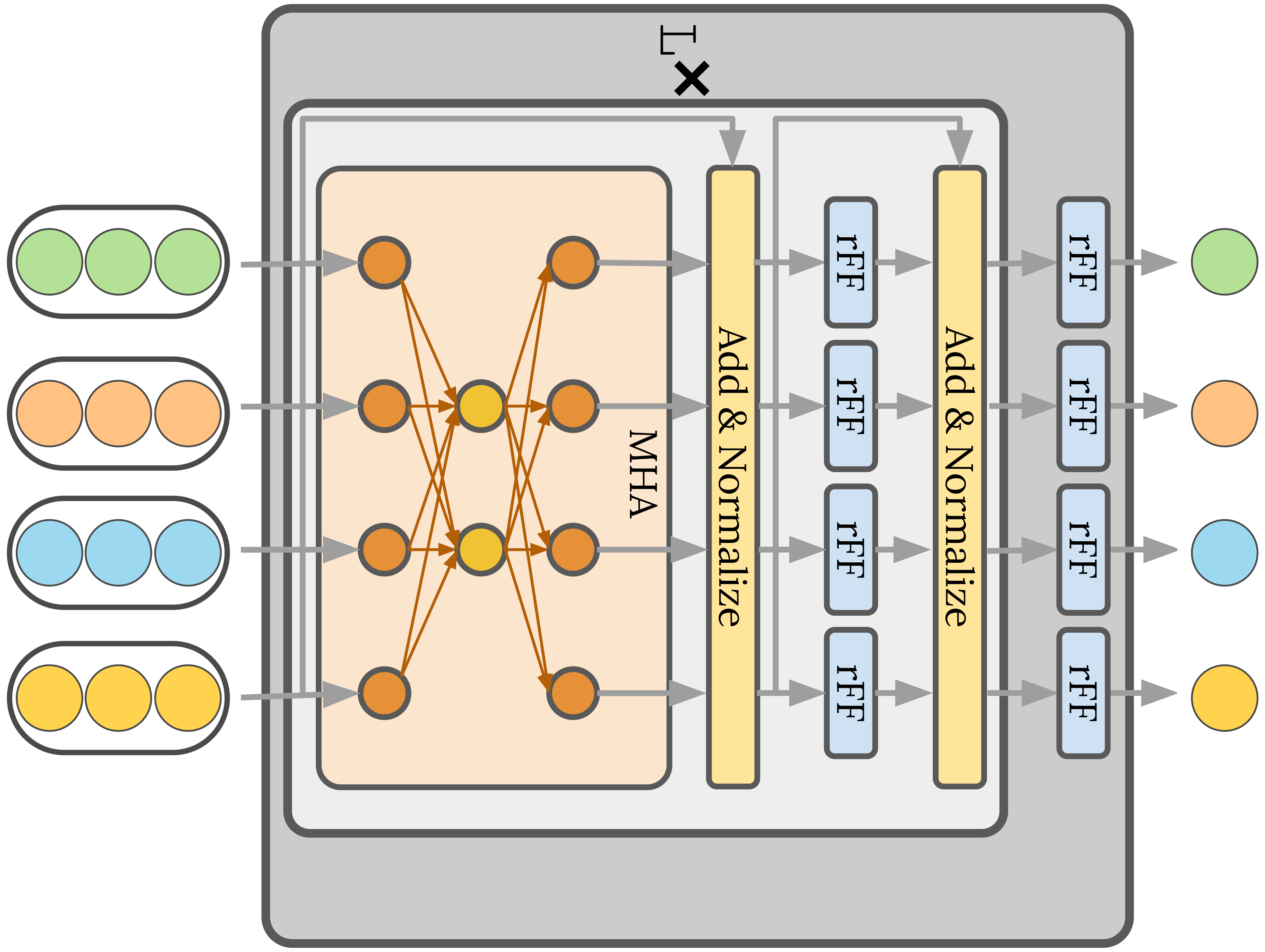


Steps:

1. **Induced self-attention block**
   Multi-head self-attention via *K* induced points

Mostra: Encoder

Set Transformer encoder giving contextualised features for each track in the pool through L layers

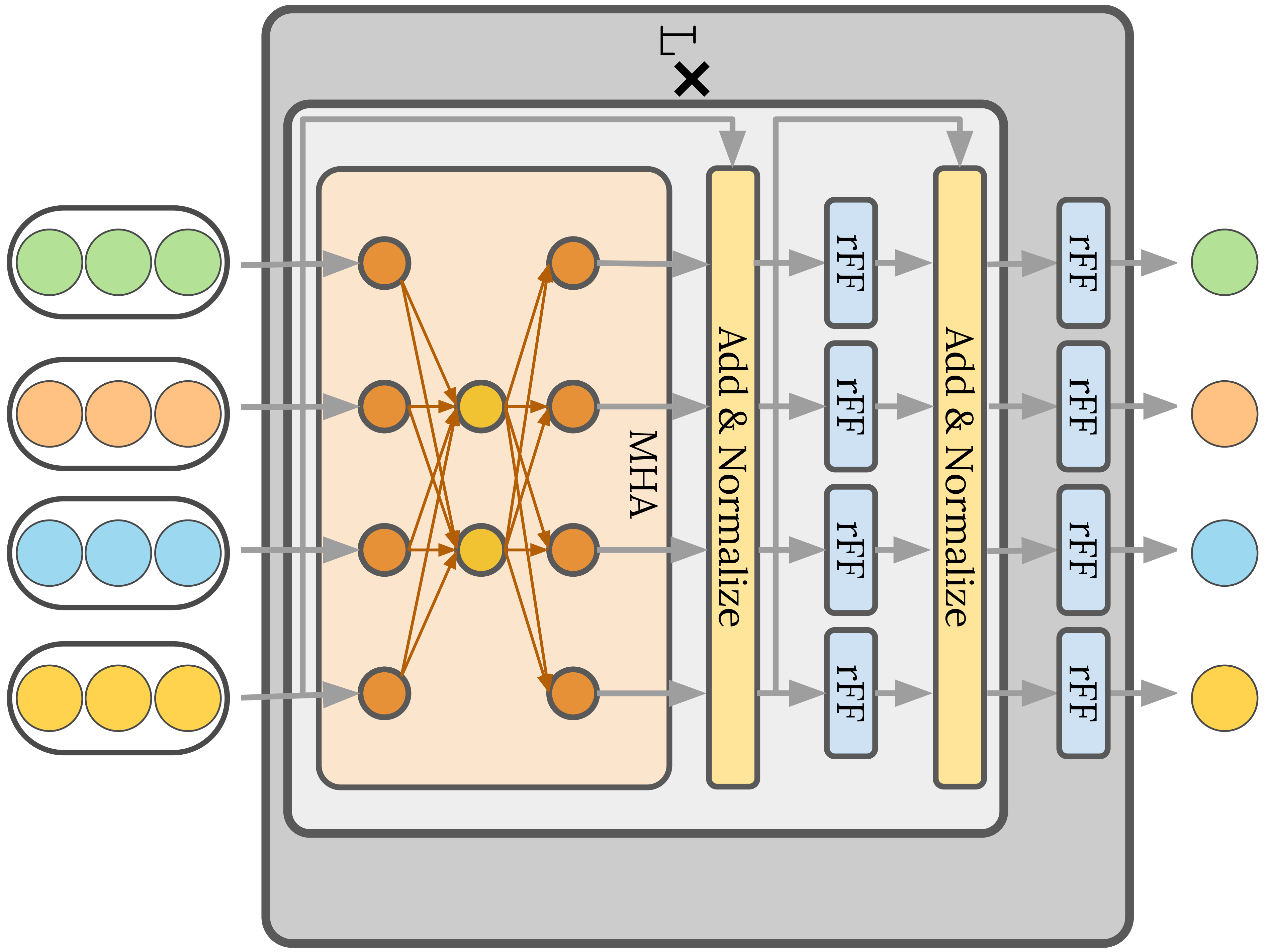


Steps:

1. **Induced self-attention block**
   Multi-head self-attention via *K* induced points
2. **Feed-forward block**
   2-layer MLP projecting each track's features

Mostra: Encoder

Set Transformer encoder giving contextualised features for each track in the pool through L layers

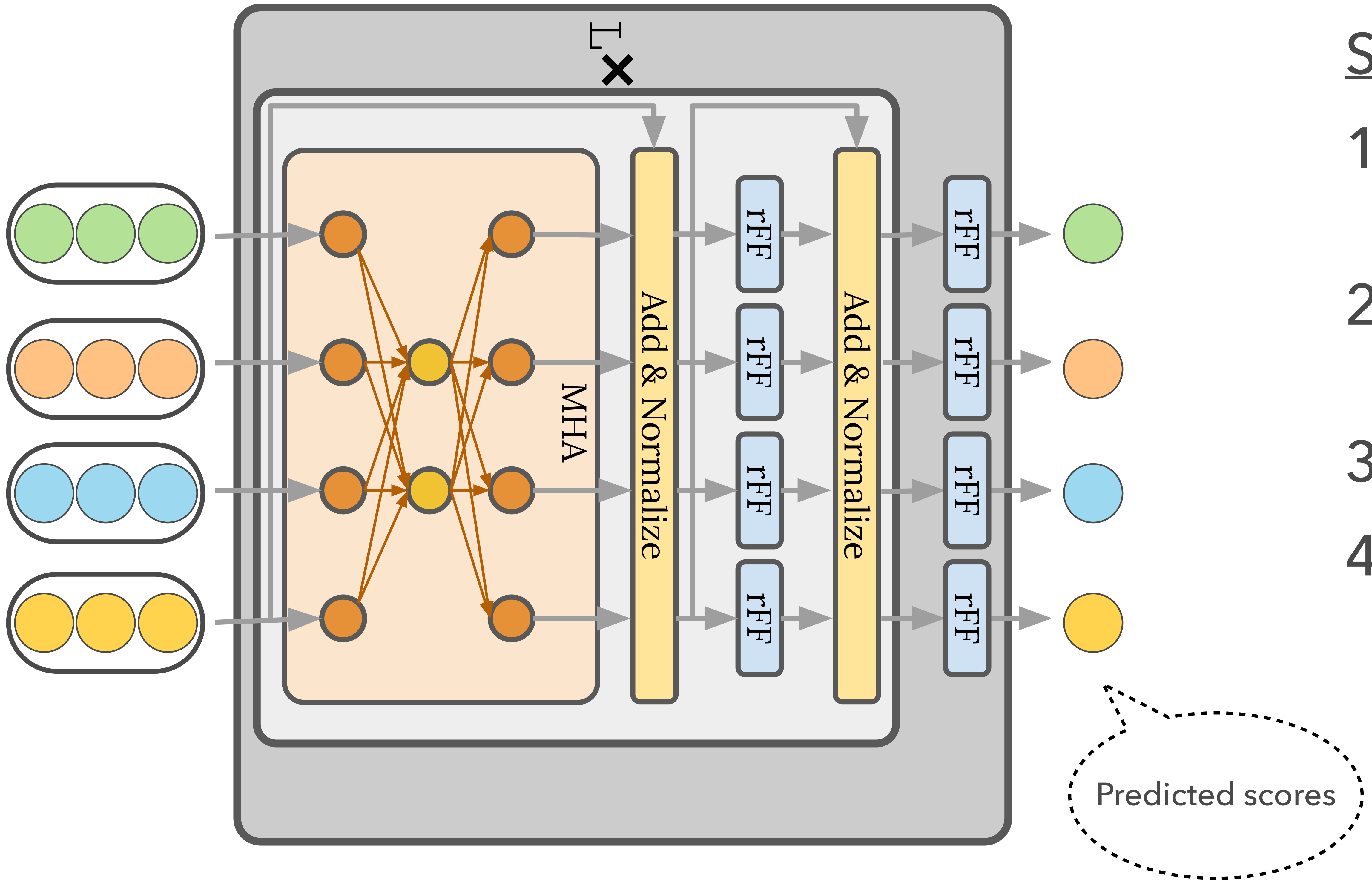


Steps:

1. **Induced self-attention block**
   Multi-head self-attention via *K* induced points
2. **Feed-forward block**
   2-layer MLP projecting each track's features
3. Repeat for L layers

Mostra: Encoder

Set Transformer encoder giving contextualised features for each track in the pool through L layers

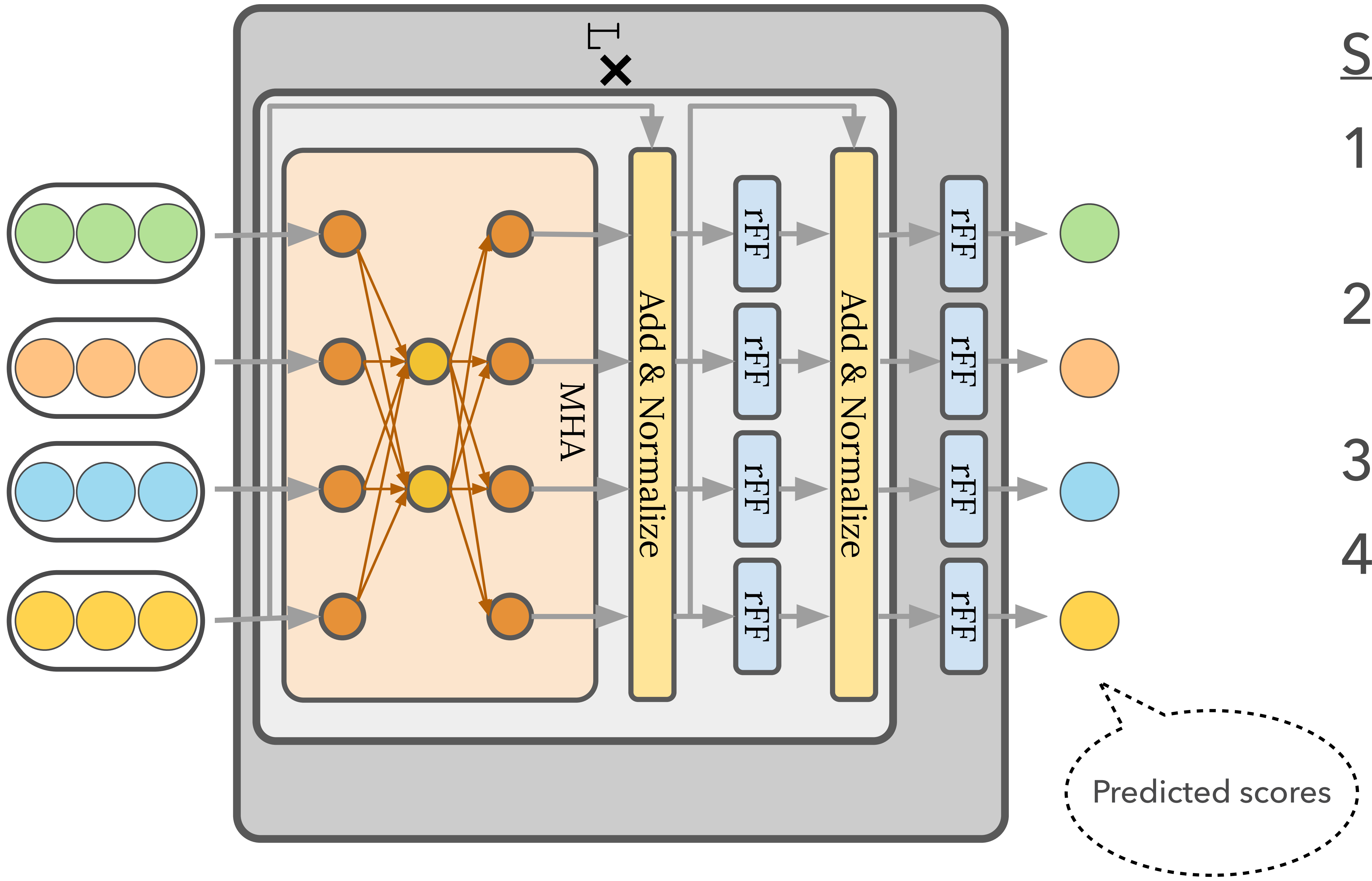


Steps:

1. **Induced self-attention block**
   Multi-head self-attention via *K* induced points
2. **Feed-forward block**
   2-layer MLP projecting each track's features
3. Repeat for L layers
4. **Mapping layer**
   Project features into "*set-aware*" relevance scores

Mostra: Encoder

Set Transformer encoder giving contextualised features for each track in the pool through L layers

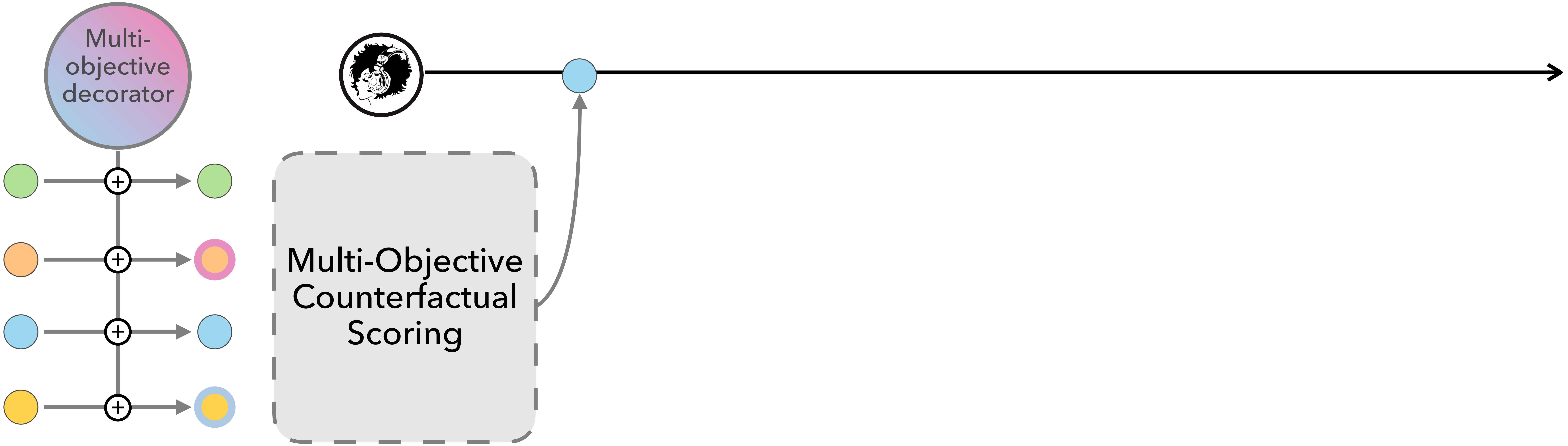


Steps:

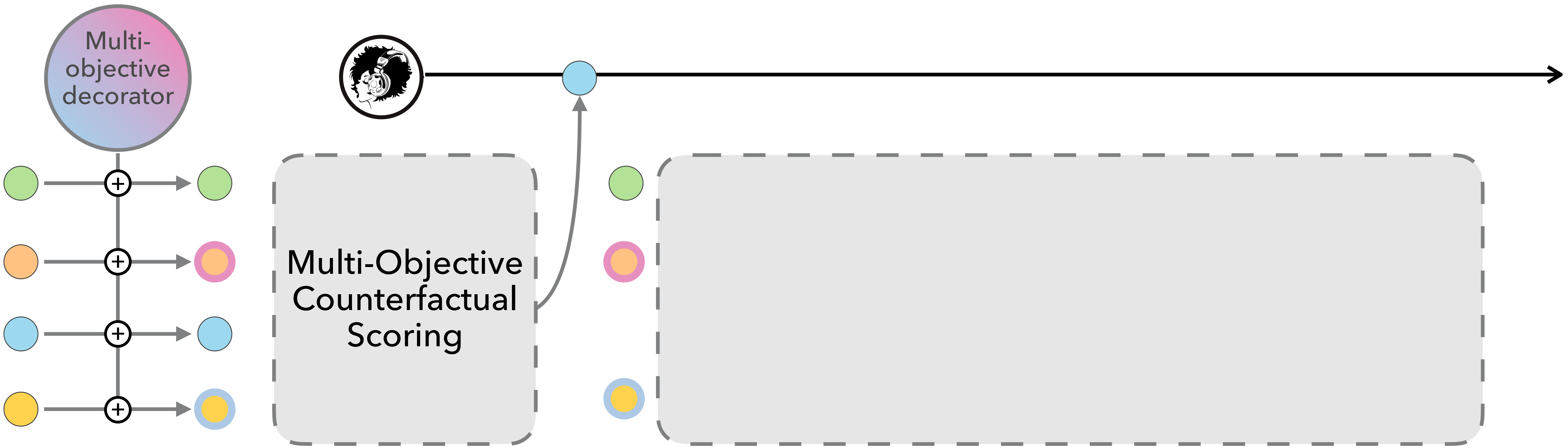
1. **Induced self-attention block**
   Multi-head self-attention via *K* induced points
2. **Feed-forward block**
   2-layer MLP projecting each track's features
3. Repeat for L layers
4. **Mapping layer**
   Project features into "*set-aware*" relevance scores

Set Transformer
- Does not model position ➡ permutation-invariant representations
- Induced self-attention scales to large input sets for K ≪ N

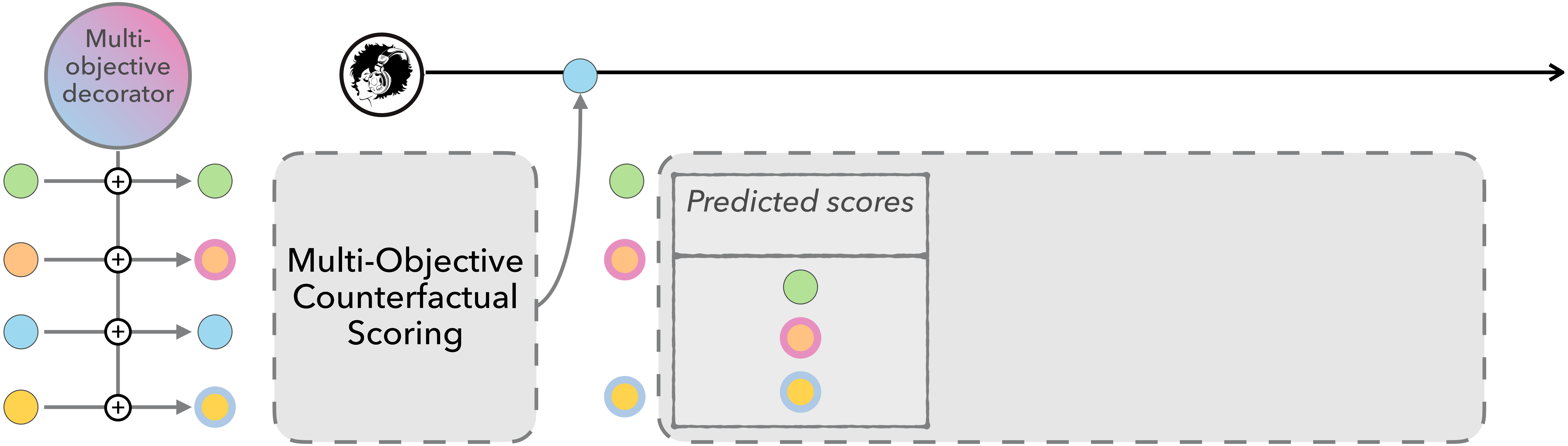
Mostra: Multi-Objective Counterfactual Beam Search



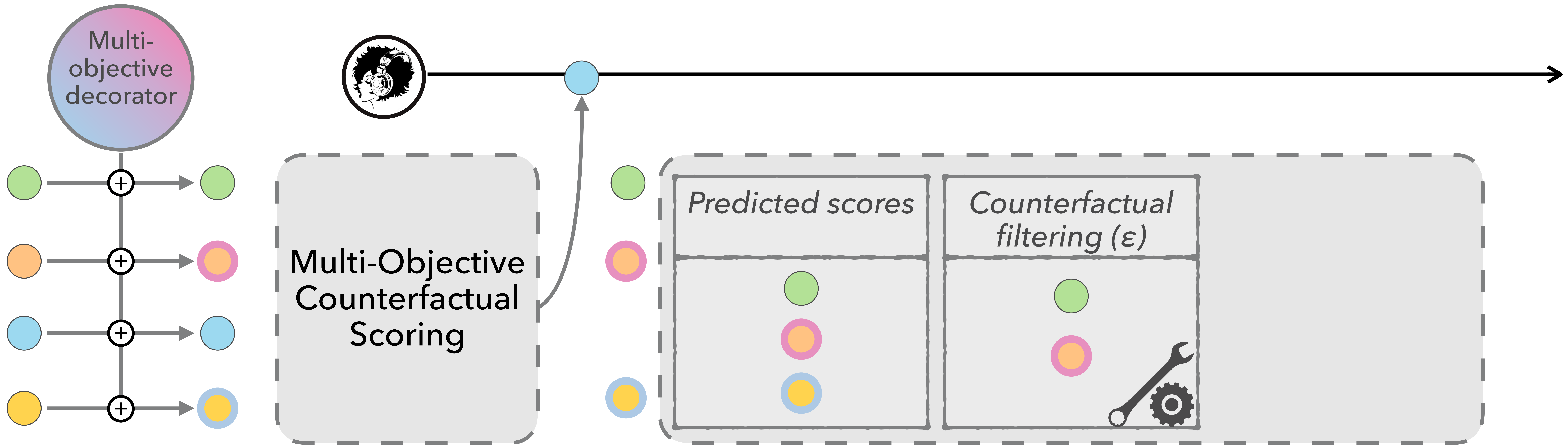
Mostra: Multi-Objective Counterfactual Beam Search



Mostra: Multi-Objective Counterfactual Beam Search

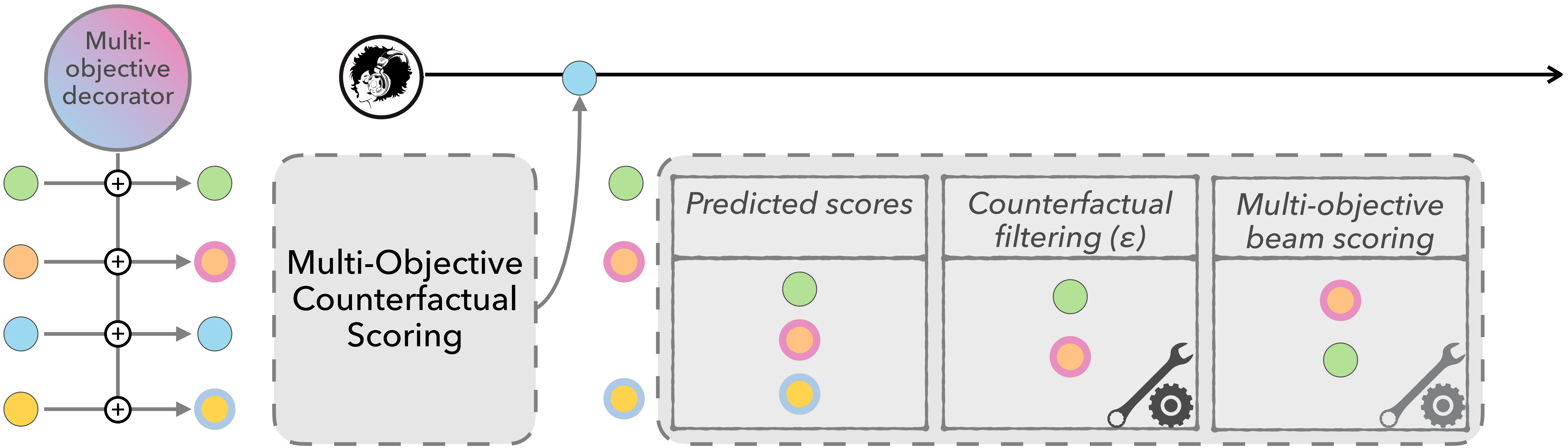


Mostra: Multi-Objective Counterfactual Beam Search



Counterfactual scoring
Filter out tracks whose predicted SAT scores are ≥ε points away from the next best track
➡ limits potential losses wrt training metric (SAT)

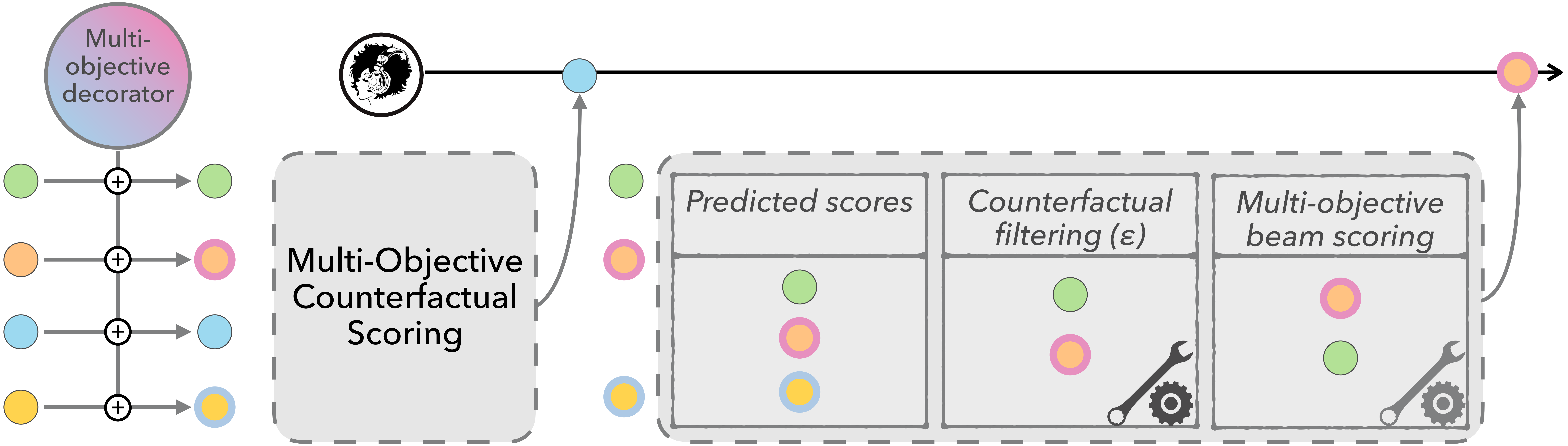
Mostra: Multi-Objective Counterfactual Beam Search



Counterfactual scoring
Filter out tracks whose predicted SAT scores are ≥ε points away from the next best track
➡ limits potential losses wrt training metric (SAT)

Submodular Multi-Objective Beam Scoring
Encourage selection of tracks from different objectives
➡ tracks from objectives not well represented in the beam so far are scored higher

Mostra: Multi-Objective Counterfactual Beam Search



Counterfactual scoring
Filter out tracks whose predicted SAT scores are ≥ε points away from the next best track
➡ limits potential losses wrt training metric (SAT)

Submodular Multi-Objective Beam Scoring
Encourage selection of tracks from different objectives
➡ tracks from objectives not well represented in the beam so far are scored higher

Multi-Objective Counterfactual Beam Search Algorithm

Algorithm 1: Multi-objective counterfactual beam search

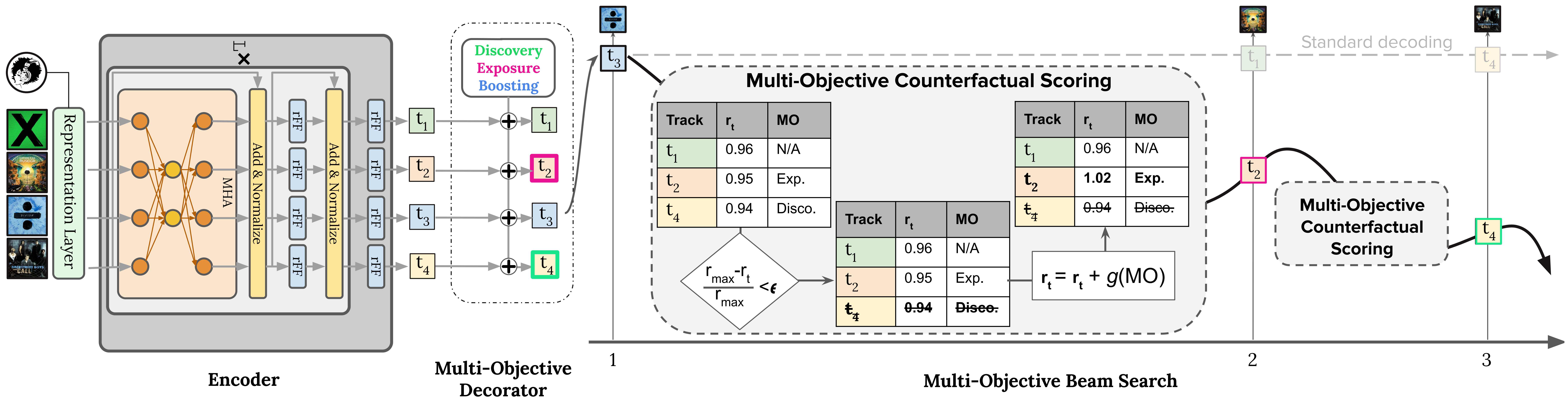
inputs: $\mathbf{t}$: tracks $\langle \mathbf{t}_1, \ldots, \mathbf{t}_N \rangle$, $\mathbf{r}$: relevance scores $\langle \mathbf{r}_1, \ldots, \mathbf{r}_N \rangle$, $\mathbf{E}$: exposure scores $\langle \langle \mathbf{E}_{11}, \ldots, \mathbf{E}_{1J} \rangle, \ldots, \langle \mathbf{E}_{N1}, \ldots, \mathbf{E}_{NJ} \rangle$, N: set length, J: number of exposure objectives, k: maximum beam size

$B_0 \leftarrow \{\langle 0, \emptyset \rangle\}$
for $n \leftarrow 1$ **to** N **do**
    $B \leftarrow \emptyset$
    for $\langle s, \mathbf{y} \rangle \in B_{n-1}$ **do**
        $\mathbf{a} \leftarrow \mathbf{t} \setminus \mathbf{y}$
        for $t \in \mathbf{a}$ **do**
            $s^c \leftarrow$ `CounterfactualScore` $(\mathbf{r}_t, \mathbf{a})$ ▷ Eq. (11)
            $s^s \leftarrow$ `SequentialScore` (s, s^c)
            $s^e \leftarrow$ `MOScore` $([\mathbf{y} \parallel t], \mathbf{E}, s^s, n)$ ▷ Eq. (13)
            B.add$(\langle s^e, [\mathbf{y} \parallel t] \rangle)$
        end
    end
    $B_n \leftarrow B$.top(k)
end
return B_N.max()

Mostra: At a Glance

🤖 **Mostra: M**ulti-**o**bjective **S**et **Tra**nsformer

Transformer model for flexible recommendations through multi-objective counterfactual decoding



Performance

Experimental Setup

🗂️ **Data**

- Spotify radio-like music streaming sessions

🧠 **Baselines**

- Single-objective item-level models
  - ▶ Relevance ranker
  - ▶ DNN
- Single-objective set-level models
  - ▶ SetRank (Pang et al., SIGIR'20)
  - ▶ Encoder–decoder models (e.g. SetTransformer–GRU)
- Multi-objective set-level models
  - ▶ MO-LTR
  - ▶ Mostra-WtSum
  - ▶ Mostra

Results

Method	NDCG@5 [%] SAT	Boost	Expos.	Disco.	NDCG@10 [%] SAT	Boost	Expos.	Disco.
Track-level Single-objective								
Relevance ranker	63.95	28.01	51.90	6.03	71.66	34.53	58.34	9.36
$\text{DNN}_{\text{AttRank}}$	69.14	27.28	50.03	17.29	76.16	33.84	56.93	23.08
DNN_{BCE}	69.48	28.68	45.35	15.87	75.90	35.10	53.27	21.67
DNN_{RMSE}	69.54	28.35	45.06	15.80	75.90	34.79	52.97	21.63
Set-level Single-objective								
$\text{SetTRM2GRU}_{\text{AttRank}}$	68.19	27.68	51.25	21.51	74.88	34.23	57.82	25.90
$\text{SetTRM2TRM}_{\text{AttRank}}$	65.43	27.45	50.75	19.62	72.88	34.05	57.52	24.91
SetRank	68.78	28.95	45.64	17.04	75.37	35.28	53.50	22.61
$\text{SetRank}_{\text{BCE}}$	65.47	26.67	52.56	20.59	72.79	33.52	59.03	25.55
$\text{SetRank}_{\text{RMSE}}$	70.11	28.94	44.81	15.48	76.37	35.21	52.87	21.39
Set-level Learnt Multi-objective								
$\text{MO-LTR}_{\text{AttRank}}$	54.63	27.19	54.12	48.76	64.70	33.87	60.26	48.76
$\text{MO-LTR}_{\text{RMSE}}$	53.74	40.42	70.09	48.77	64.07	44.75	72.94	48.77
Set-level Online Multi-objective								
Mostra-WtSum ($\alpha = 1.0$)	61.05	52.03	70.58	34.61	69.47	53.81	72.92	36.75
Mostra-WtSum ($\alpha = 0.5$)	61.19	51.86	70.37	34.31	69.58	53.67	72.76	36.50
Mostra-WtSum ($\alpha = 0.1$)	64.91	45.16	61.45	26.91	72.40	48.20	65.59	30.48
Mostra-WtSum ($\alpha = 0.01$)	69.49	31.43	47.18	16.92	75.88	37.18	54.62	22.50
Mostra ($\epsilon = 0.01$)	$69.56^{\Uparrow}$	$31.08^{\Uparrow}$	$46.56^{\Uparrow}$	16.77	$75.94^{\Uparrow}$	$36.90^{\Uparrow}$	$54.17^{\Uparrow}$	$22.37^{\downarrow}$
Mostra ($\epsilon = 0.05$)	$67.66^{\Downarrow}$	$36.76^{\Uparrow}$	$51.14^{\Uparrow}$	$21.13^{\Uparrow}$	$74.60^{\Uparrow}$	$41.09^{\Uparrow}$	$57.39^{\Uparrow}$	$25.48^{\Uparrow}$
Mostra ($\epsilon = 0.10$)	$66.33^{\Downarrow}$	$39.83^{\Uparrow}$	$53.86^{\Uparrow}$	$24.00^{\Uparrow}$	$73.54^{\Uparrow}$	$43.74^{\Uparrow}$	$59.65^{\Uparrow}$	$27.85^{\Uparrow}$

Results

Method	NDCG@5 [%] SAT	Boost	Expos.	Disco.	NDCG@10 [%] SAT	Boost	Expos.	Disco.
Track-level Single-objective								
Relevance ranker	63.95	28.01	51.90	6.03	71.66	34.53	58.34	9.36
$\text{DNN}_{\text{AttRank}}$	69.14	27.28	50.03	17.29	76.16	33.84	56.93	23.08
DNN_{BCE}	69.48	28.68	45.35	15.87	75.90	35.10	53.27	21.67
DNN_{RMSE}	69.54	28.35	45.06	15.80	75.90	34.79	52.97	21.63
Set-level Single-objective								
$\text{SetTRM2GRU}_{\text{AttRank}}$	68.19	27.68	51.25	21.51	74.88	34.23	57.82	25.90
$\text{SetTRM2TRM}_{\text{AttRank}}$	65.43	27.45	50.75	19.62	72.88	34.05	57.52	24.91
SetRank	68.78	28.95	45.64	17.04	75.37	35.28	53.50	22.61
$\text{SetRank}_{\text{BCE}}$	65.47	26.67	52.56	20.59	72.79	33.52	59.03	25.55
$\text{SetRank}_{\text{RMSE}}$	70.11	28.94	44.81	15.48	76.37	35.21	52.87	21.39
Set-level Learnt Multi-objective								
$\text{MO-LTR}_{\text{AttRank}}$	54.63	27.19	54.12	48.76	64.70	33.87	60.26	48.76
$\text{MO-LTR}_{\text{RMSE}}$	53.74	40.42	70.09	48.77	64.07	44.75	72.94	48.77
Set-level Online Multi-objective								
Mostra-WtSum ($\alpha = 1.0$)	61.05	52.03	70.58	34.61	69.47	53.81	72.92	36.75
Mostra-WtSum ($\alpha = 0.5$)	61.19	51.86	70.37	34.31	69.58	53.67	72.76	36.50
Mostra-WtSum ($\alpha = 0.1$)	64.91	45.16	61.45	26.91	72.40	48.20	65.59	30.48
Mostra-WtSum ($\alpha = 0.01$)	69.49	31.43	47.18	16.92	75.88	37.18	54.62	22.50
Mostra ($\epsilon = 0.01$)	$69.56^{\Uparrow}$	$31.08^{\Uparrow}$	$46.56^{\Uparrow}$	16.77	$75.94^{\Uparrow}$	$36.90^{\Uparrow}$	$54.17^{\Uparrow}$	$22.37^{\downarrow}$
Mostra ($\epsilon = 0.05$)	$67.66^{\Downarrow}$	$36.76^{\Uparrow}$	$51.14^{\Uparrow}$	$21.13^{\Uparrow}$	$74.60^{\Uparrow}$	$41.09^{\Uparrow}$	$57.39^{\Uparrow}$	$25.48^{\Uparrow}$
Mostra ($\epsilon = 0.10$)	$66.33^{\Downarrow}$	$39.83^{\Uparrow}$	$53.86^{\Uparrow}$	$24.00^{\Uparrow}$	$73.54^{\Uparrow}$	$43.74^{\Uparrow}$	$59.65^{\Uparrow}$	$27.85^{\Uparrow}$

Track-level VS. set-level models

- On par
- Set-level methods are more efficient

Results

Method	NDCG@5 [%] SAT	Boost	Expos.	Disco.	NDCG@10 [%] SAT	Boost	Expos.	Disco.
Track-level Single-objective								
Relevance ranker	63.95	28.01	51.90	6.03	71.66	34.53	58.34	9.36
$\text{DNN}_{\text{AttRank}}$	69.14	27.28	50.03	17.29	76.16	33.84	56.93	23.08
DNN_{BCE}	69.48	28.68	45.35	15.87	75.90	35.10	53.27	21.67
DNN_{RMSE}	69.54	28.35	45.06	15.80	75.90	34.79	52.97	21.63
Set-level Single-objective								
$\text{SetTRM2GRU}_{\text{AttRank}}$	68.19	27.68	51.25	21.51	74.88	34.23	57.82	25.90
$\text{SetTRM2TRM}_{\text{AttRank}}$	65.43	27.45	50.75	19.62	72.88	34.05	57.52	24.91
SetRank	68.78	28.95	45.64	17.04	75.37	35.28	53.50	22.61
$\text{SetRank}_{\text{BCE}}$	65.47	26.67	52.56	20.59	72.79	33.52	59.03	25.55
$\text{SetRank}_{\text{RMSE}}$	70.11	28.94	44.81	15.48	76.37	35.21	52.87	21.39
Set-level Learnt Multi-objective								
$\text{MO-LTR}_{\text{AttRank}}$	54.63	27.19	54.12	48.76	64.70	33.87	60.26	48.76
$\text{MO-LTR}_{\text{RMSE}}$	53.74	40.42	70.09	48.77	64.07	44.75	72.94	48.77
Set-level Online Multi-objective								
Mostra-WtSum ($\alpha = 1.0$)	61.05	52.03	70.58	34.61	69.47	53.81	72.92	36.75
Mostra-WtSum ($\alpha = 0.5$)	61.19	51.86	70.37	34.31	69.58	53.67	72.76	36.50
Mostra-WtSum ($\alpha = 0.1$)	64.91	45.16	61.45	26.91	72.40	48.20	65.59	30.48
Mostra-WtSum ($\alpha = 0.01$)	69.49	31.43	47.18	16.92	75.88	37.18	54.62	22.50
Mostra ($\epsilon = 0.01$)	$69.56^{\Uparrow}$	$31.08^{\Uparrow}$	$46.56^{\Uparrow}$	16.77	$75.94^{\Uparrow}$	$36.90^{\Uparrow}$	$54.17^{\Uparrow}$	$22.37^{\downarrow}$
Mostra ($\epsilon = 0.05$)	$67.66^{\Downarrow}$	$36.76^{\Uparrow}$	$51.14^{\Uparrow}$	$21.13^{\Uparrow}$	$74.60^{\Uparrow}$	$41.09^{\Uparrow}$	$57.39^{\Uparrow}$	$25.48^{\Uparrow}$
Mostra ($\epsilon = 0.10$)	$66.33^{\Downarrow}$	$39.83^{\Uparrow}$	$53.86^{\Uparrow}$	$24.00^{\Uparrow}$	$73.54^{\Uparrow}$	$43.74^{\Uparrow}$	$59.65^{\Uparrow}$	$27.85^{\Uparrow}$

Track-level VS. set-level models

- On par
- Set-level methods are more efficient

Loss function

- Point-wise (RMSE) > list-wise (AttRank)

Results

Method	NDCG@5 [%] SAT	Boost	Expos.	Disco.	NDCG@10 [%] SAT	Boost	Expos.	Disco.
Track-level Single-objective								
Relevance ranker	63.95	28.01	51.90	6.03	71.66	34.53	58.34	9.36
$\text{DNN}_{\text{AttRank}}$	69.14	27.28	50.03	17.29	76.16	33.84	56.93	23.08
DNN_{BCE}	69.48	28.68	45.35	15.87	75.90	35.10	53.27	21.67
DNN_{RMSE}	69.54	28.35	45.06	15.80	75.90	34.79	52.97	21.63
Set-level Single-objective								
$\text{SetTRM2GRU}_{\text{AttRank}}$	68.19	27.68	51.25	21.51	74.88	34.23	57.82	25.90
$\text{SetTRM2TRM}_{\text{AttRank}}$	65.43	27.45	50.75	19.62	72.88	34.05	57.52	24.91
SetRank	68.78	28.95	45.64	17.04	75.37	35.28	53.50	22.61
$\text{SetRank}_{\text{BCE}}$	65.47	26.67	52.56	20.59	72.79	33.52	59.03	25.55
$\text{SetRank}_{\text{RMSE}}$	70.11	28.94	44.81	15.48	76.37	35.21	52.87	21.39
Set-level Learnt Multi-objective								
$\text{MO-LTR}_{\text{AttRank}}$	54.63	27.19	54.12	48.76	64.70	33.87	60.26	48.76
$\text{MO-LTR}_{\text{RMSE}}$	53.74	40.42	70.09	48.77	64.07	44.75	72.94	48.77
Set-level Online Multi-objective								
Mostra-WtSum ($\alpha = 1.0$)	61.05	52.03	70.58	34.61	69.47	53.81	72.92	36.75
Mostra-WtSum ($\alpha = 0.5$)	61.19	51.86	70.37	34.31	69.58	53.67	72.76	36.50
Mostra-WtSum ($\alpha = 0.1$)	64.91	45.16	61.45	26.91	72.40	48.20	65.59	30.48
Mostra-WtSum ($\alpha = 0.01$)	69.49	31.43	47.18	16.92	75.88	37.18	54.62	22.50
Mostra ($\epsilon = 0.01$)	$69.56^{\Uparrow}$	$31.08^{\Uparrow}$	$46.56^{\Uparrow}$	16.77	$75.94^{\Uparrow}$	$36.90^{\Uparrow}$	$54.17^{\Uparrow}$	$22.37^{\downarrow}$
Mostra ($\epsilon = 0.05$)	$67.66^{\Downarrow}$	$36.76^{\Uparrow}$	$51.14^{\Uparrow}$	$21.13^{\Uparrow}$	$74.60^{\Uparrow}$	$41.09^{\Uparrow}$	$57.39^{\Uparrow}$	$25.48^{\Uparrow}$
Mostra ($\epsilon = 0.10$)	$66.33^{\Downarrow}$	$39.83^{\Uparrow}$	$53.86^{\Uparrow}$	$24.00^{\Uparrow}$	$73.54^{\Uparrow}$	$43.74^{\Uparrow}$	$59.65^{\Uparrow}$	$27.85^{\Uparrow}$

Track-level VS. set-level models

- On par
- Set-level methods are more efficient

Loss function

- Point-wise (RMSE) > list-wise (AttRank)

Multi-Objective models

Results

Method	NDCG@5 [%] SAT	NDCG@5 [%] Boost	NDCG@5 [%] Expos.	NDCG@5 [%] Disco.	NDCG@10 [%] SAT	NDCG@10 [%] Boost	NDCG@10 [%] Expos.	NDCG@10 [%] Disco.
Track-level Single-objective								
Relevance ranker	63.95	28.01	51.90	6.03	71.66	34.53	58.34	9.36
$\text{DNN}_{\text{AttRank}}$	69.14	27.28	50.03	17.29	76.16	33.84	56.93	23.08
DNN_{BCE}	69.48	28.68	45.35	15.87	75.90	35.10	53.27	21.67
DNN_{RMSE}	69.54	28.35	45.06	15.80	75.90	34.79	52.97	21.63
Set-level Single-objective								
$\text{SetTRM2GRU}_{\text{AttRank}}$	68.19	27.68	51.25	21.51	74.88	34.23	57.82	25.90
$\text{SetTRM2TRM}_{\text{AttRank}}$	65.43	27.45	50.75	19.62	72.88	34.05	57.52	24.91
SetRank	68.78	28.95	45.64	17.04	75.37	35.28	53.50	22.61
$\text{SetRank}_{\text{BCE}}$	65.47	26.67	52.56	20.59	72.79	33.52	59.03	25.55
$\text{SetRank}_{\text{RMSE}}$	70.11	28.94	44.81	15.48	76.37	35.21	52.87	21.39
Set-level Learnt Multi-objective								
$\text{MO-LTR}_{\text{AttRank}}$	54.63	27.19	54.12	48.76	64.70	33.87	60.26	48.76
$\text{MO-LTR}_{\text{RMSE}}$	53.74	40.42	70.09	48.77	64.07	44.75	72.94	48.77
Set-level Online Multi-objective								
Mostra-WtSum ($\alpha = 1.0$)	61.05	52.03	70.58	34.61	69.47	53.81	72.92	36.75
Mostra-WtSum ($\alpha = 0.5$)	61.19	51.86	70.37	34.31	69.58	53.67	72.76	36.50
Mostra-WtSum ($\alpha = 0.1$)	64.91	45.16	61.45	26.91	72.40	48.20	65.59	30.48
Mostra-WtSum ($\alpha = 0.01$)	69.49	31.43	47.18	16.92	75.88	37.18	54.62	22.50
Mostra ($\epsilon = 0.01$)	$69.56^{\Uparrow}$	$31.08^{\Uparrow}$	$46.56^{\Uparrow}$	16.77	$75.94^{\Uparrow}$	$36.90^{\Uparrow}$	$54.17^{\Uparrow}$	$22.37^{\downarrow}$
Mostra ($\epsilon = 0.05$)	$67.66^{\Downarrow}$	$36.76^{\Uparrow}$	$51.14^{\Uparrow}$	$21.13^{\Uparrow}$	$74.60^{\Uparrow}$	$41.09^{\Uparrow}$	$57.39^{\Uparrow}$	$25.48^{\Uparrow}$
Mostra ($\epsilon = 0.10$)	$66.33^{\Downarrow}$	$39.83^{\Uparrow}$	$53.86^{\Uparrow}$	$24.00^{\Uparrow}$	$73.54^{\Uparrow}$	$43.74^{\Uparrow}$	$59.65^{\Uparrow}$	$27.85^{\Uparrow}$

Track-level VS. set-level models

- On par
- Set-level methods are more efficient

Loss function

- Point-wise (RMSE) > list-wise (AttRank)

Multi-Objective models

- Naive LTR models don't work well

Results

Method	NDCG@5 [%] SAT	Boost	Expos.	Disco.	NDCG@10 [%] SAT	Boost	Expos.	Disco.
Track-level Single-objective								
Relevance ranker	63.95	28.01	51.90	6.03	71.66	34.53	58.34	9.36
$\text{DNN}_{\text{AttRank}}$	69.14	27.28	50.03	17.29	76.16	33.84	56.93	23.08
DNN_{BCE}	69.48	28.68	45.35	15.87	75.90	35.10	53.27	21.67
DNN_{RMSE}	69.54	28.35	45.06	15.80	75.90	34.79	52.97	21.63
Set-level Single-objective								
$\text{SetTRM2GRU}_{\text{AttRank}}$	68.19	27.68	51.25	21.51	74.88	34.23	57.82	25.90
$\text{SetTRM2TRM}_{\text{AttRank}}$	65.43	27.45	50.75	19.62	72.88	34.05	57.52	24.91
SetRank	68.78	28.95	45.64	17.04	75.37	35.28	53.50	22.61
$\text{SetRank}_{\text{BCE}}$	65.47	26.67	52.56	20.59	72.79	33.52	59.03	25.55
$\text{SetRank}_{\text{RMSE}}$	70.11	28.94	44.81	15.48	76.37	35.21	52.87	21.39
Set-level Learnt Multi-objective								
$\text{MO-LTR}_{\text{AttRank}}$	54.63	27.19	54.12	48.76	64.70	33.87	60.26	48.76
$\text{MO-LTR}_{\text{RMSE}}$	53.74	40.42	70.09	48.77	64.07	44.75	72.94	48.77
Set-level Online Multi-objective								
Mostra-WtSum ($\alpha = 1.0$)	61.05	52.03	70.58	34.61	69.47	53.81	72.92	36.75
Mostra-WtSum ($\alpha = 0.5$)	61.19	51.86	70.37	34.31	69.58	53.67	72.76	36.50
Mostra-WtSum ($\alpha = 0.1$)	64.91	45.16	61.45	26.91	72.40	48.20	65.59	30.48
Mostra-WtSum ($\alpha = 0.01$)	69.49	31.43	47.18	16.92	75.88	37.18	54.62	22.50
Mostra ($\epsilon = 0.01$)	$69.56^{\Uparrow}$	$31.08^{\Uparrow}$	$46.56^{\Uparrow}$	16.77	$75.94^{\Uparrow}$	$36.90^{\Uparrow}$	$54.17^{\Uparrow}$	$22.37^{\downarrow}$
Mostra ($\epsilon = 0.05$)	$67.66^{\Downarrow}$	$36.76^{\Uparrow}$	$51.14^{\Uparrow}$	$21.13^{\Uparrow}$	$74.60^{\Uparrow}$	$41.09^{\Uparrow}$	$57.39^{\Uparrow}$	$25.48^{\Uparrow}$
Mostra ($\epsilon = 0.10$)	$66.33^{\Downarrow}$	$39.83^{\Uparrow}$	$53.86^{\Uparrow}$	$24.00^{\Uparrow}$	$73.54^{\Uparrow}$	$43.74^{\Uparrow}$	$59.65^{\Uparrow}$	$27.85^{\Uparrow}$

Track-level VS. set-level models

- On par
- Set-level methods are more efficient

Loss function

- Point-wise (RMSE) > list-wise (AttRank)

Multi-Objective models

- Naive LTR models don't work well
- Decoding with a weighted sum performs better, but hard to control

Results

Method	NDCG@5 [%] SAT	Boost	Expos.	Disco.	NDCG@10 [%] SAT	Boost	Expos.	Disco.
Track-level Single-objective								
Relevance ranker	63.95	28.01	51.90	6.03	71.66	34.53	58.34	9.36
$\text{DNN}_{\text{AttRank}}$	69.14	27.28	50.03	17.29	76.16	33.84	56.93	23.08
DNN_{BCE}	69.48	28.68	45.35	15.87	75.90	35.10	53.27	21.67
DNN_{RMSE}	69.54	28.35	45.06	15.80	75.90	34.79	52.97	21.63
Set-level Single-objective								
$\text{SetTRM2GRU}_{\text{AttRank}}$	68.19	27.68	51.25	21.51	74.88	34.23	57.82	25.90
$\text{SetTRM2TRM}_{\text{AttRank}}$	65.43	27.45	50.75	19.62	72.88	34.05	57.52	24.91
SetRank	68.78	28.95	45.64	17.04	75.37	35.28	53.50	22.61
$\text{SetRank}_{\text{BCE}}$	65.47	26.67	52.56	20.59	72.79	33.52	59.03	25.55
$\text{SetRank}_{\text{RMSE}}$	70.11	28.94	44.81	15.48	76.37	35.21	52.87	21.39
Set-level Learnt Multi-objective								
$\text{MO-LTR}_{\text{AttRank}}$	54.63	27.19	54.12	48.76	64.70	33.87	60.26	48.76
$\text{MO-LTR}_{\text{RMSE}}$	53.74	40.42	70.09	48.77	64.07	44.75	72.94	48.77
Set-level Online Multi-objective								
Mostra-WtSum ($\alpha = 1.0$)	61.05	52.03	70.58	34.61	69.47	53.81	72.92	36.75
Mostra-WtSum ($\alpha = 0.5$)	61.19	51.86	70.37	34.31	69.58	53.67	72.76	36.50
Mostra-WtSum ($\alpha = 0.1$)	64.91	45.16	61.45	26.91	72.40	48.20	65.59	30.48
Mostra-WtSum ($\alpha = 0.01$)	69.49	31.43	47.18	16.92	75.88	37.18	54.62	22.50
Mostra ($\epsilon = 0.01$)	$69.56^{\Uparrow}$	$31.08^{\Uparrow}$	$46.56^{\Uparrow}$	16.77	$75.94^{\Uparrow}$	$36.90^{\Uparrow}$	$54.17^{\Uparrow}$	$22.37^{\downarrow}$
Mostra ($\epsilon = 0.05$)	$67.66^{\Downarrow}$	$36.76^{\Uparrow}$	$51.14^{\Uparrow}$	$21.13^{\Uparrow}$	$74.60^{\Uparrow}$	$41.09^{\Uparrow}$	$57.39^{\Uparrow}$	$25.48^{\Uparrow}$
Mostra ($\epsilon = 0.10$)	$66.33^{\Downarrow}$	$39.83^{\Uparrow}$	$53.86^{\Uparrow}$	$24.00^{\Uparrow}$	$73.54^{\Uparrow}$	$43.74^{\Uparrow}$	$59.65^{\Uparrow}$	$27.85^{\Uparrow}$

Track-level VS. set-level models

- On par
- Set-level methods are more efficient

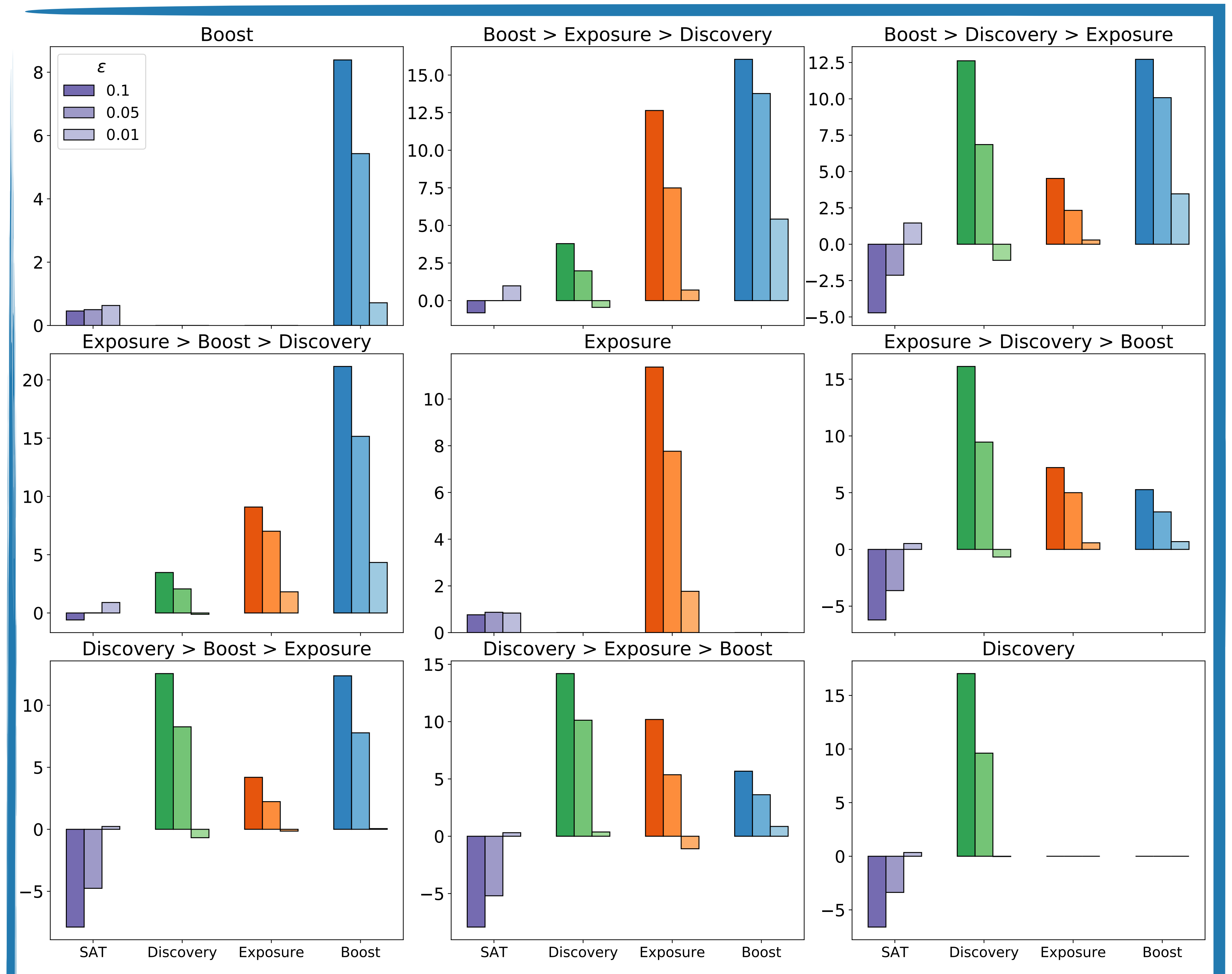
Loss function

- Point-wise (RMSE) > list-wise (AttRank)

Multi-Objective models

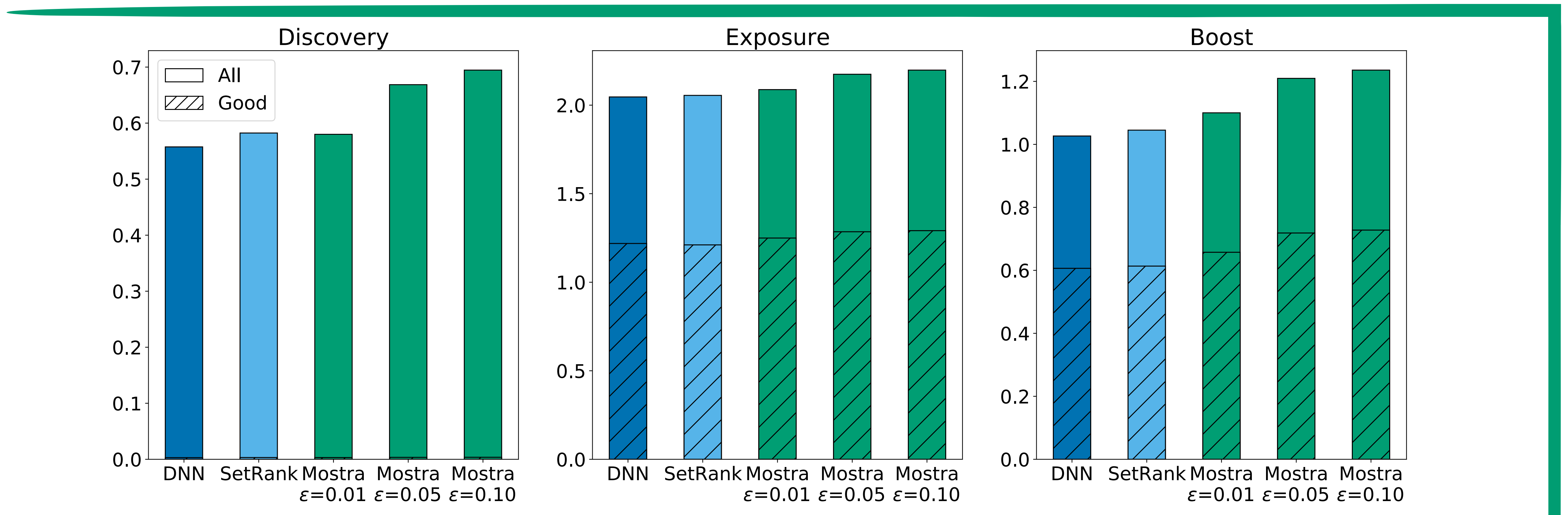
- Naive LTR models don't work well
- Decoding with a weighted sum performs better, but hard to control
- Mostra (counterfactual + submodular scoring) affords finer control

Analyses



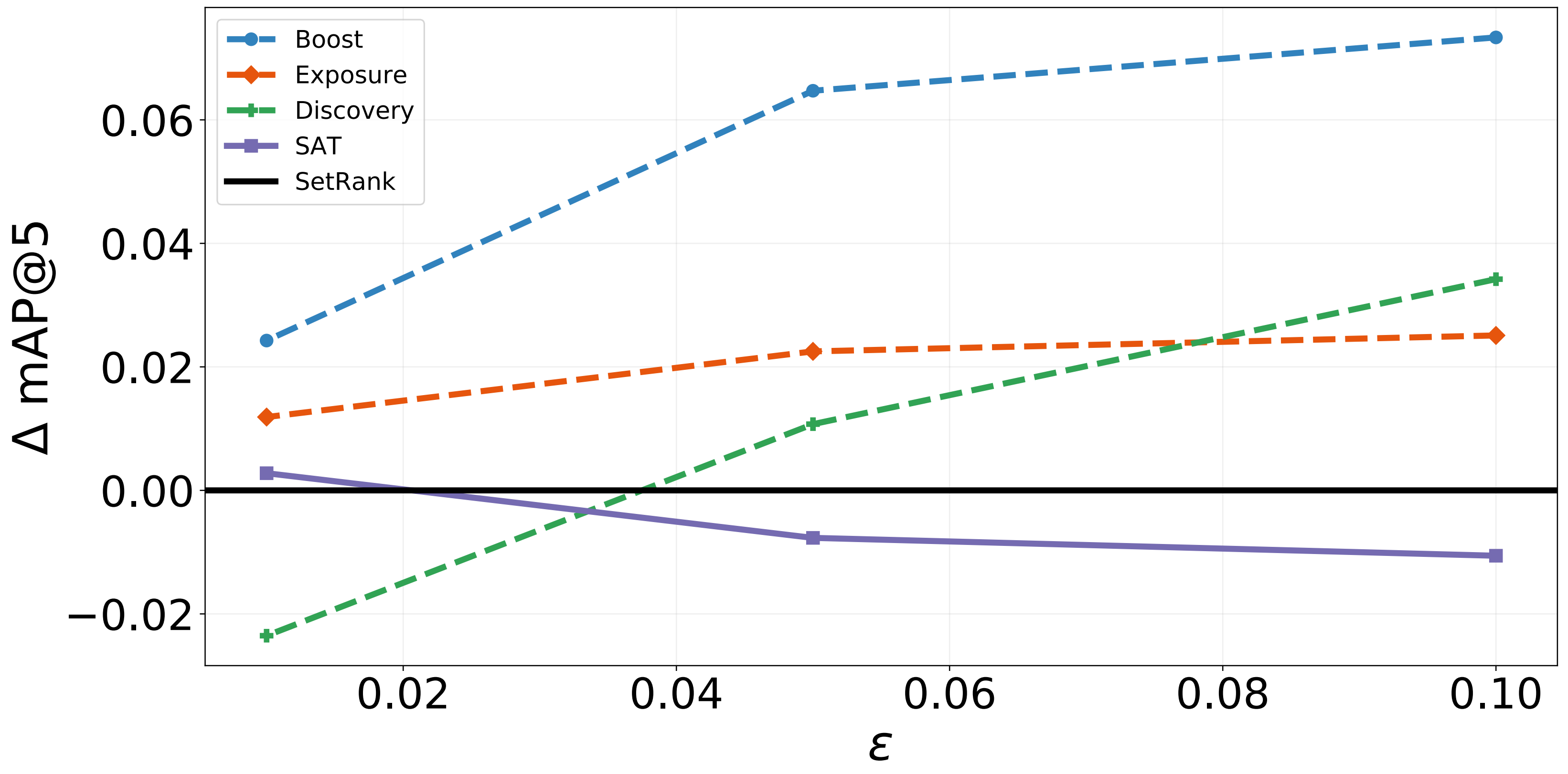
Performance by competition within sessions

- *Boost & Exposure*: Increase with SAT
- *Discovery*: Decreases SAT



Are we just showing more, or are they getting streamed as well?

- Mostra surfaces more tracks across all creator-centric objectives
- The number of streamed tracks increases too



Relaxing SAT constraints

- As we increase ε, creator-centric objectives increase
- Important: We can control this more than before

Takeaways

Conclusion

🎳 **Multi-Objective Recommendations**

- Typical of multi-stakeholder digital platforms
- Analysis in music recommendations revealed complex interplay among objectives

🏃‍♀️ **Just-in-time Multi-Objective Optimisation**

- **Mostra:** end-to-end method for scalable and flexible multi-objective recommendations
  - Flexibility given by a novel beam search algorithm
  - Significant gains over creator-centric objectives wrt SOTA with no drop in user metric

🔮 **Next:** extensions to cross-session objectives, trainable systems for causal decisioning

👩‍💻 Code is available online: github.com/spotify-research/mostra

Conclusion

Thank you

🎳 **Multi-Objective Recommendations**

- Typical of multi-stakeholder digital platforms
- Analysis in music recommendations revealed complex interplay among objectives

🏃‍♀️ **Just-in-time Multi-Objective Optimisation**

- **Mostra:** end-to-end method for scalable and flexible multi-objective recommendations
  - Flexibility given by a novel beam search algorithm
  - Significant gains over creator-centric objectives wrt SOTA with no drop in user metric

🔮 **Next:** extensions to cross-session objectives, trainable systems for causal decisioning

👩‍💻 Code is available online: github.com/spotify-research/mostra