Neural Instant Search for Music and Podcast

Speaker: Zi-Xin, Chen

Advisor: Jia-Ling, Koh

Date: 2022/05/10

Source: KDD '21

Introduction

Motivation

- In recent years, more audio streaming services are now expanding their item catalogs to support both **music and podcast**.
- The need to develop information access systems that enable efficient and effective discovery from a **heterogeneous** collection of music and podcasts is more important than ever.

Goal

- highlight the differences between query characteristics, user effort, consumption patterns, and search goals for music and podcast
- develop an **instant search** engine with **character-level** transformer-based attention model that re-ranks a set of candidate items (music and podcasts) in response to a given query prefix

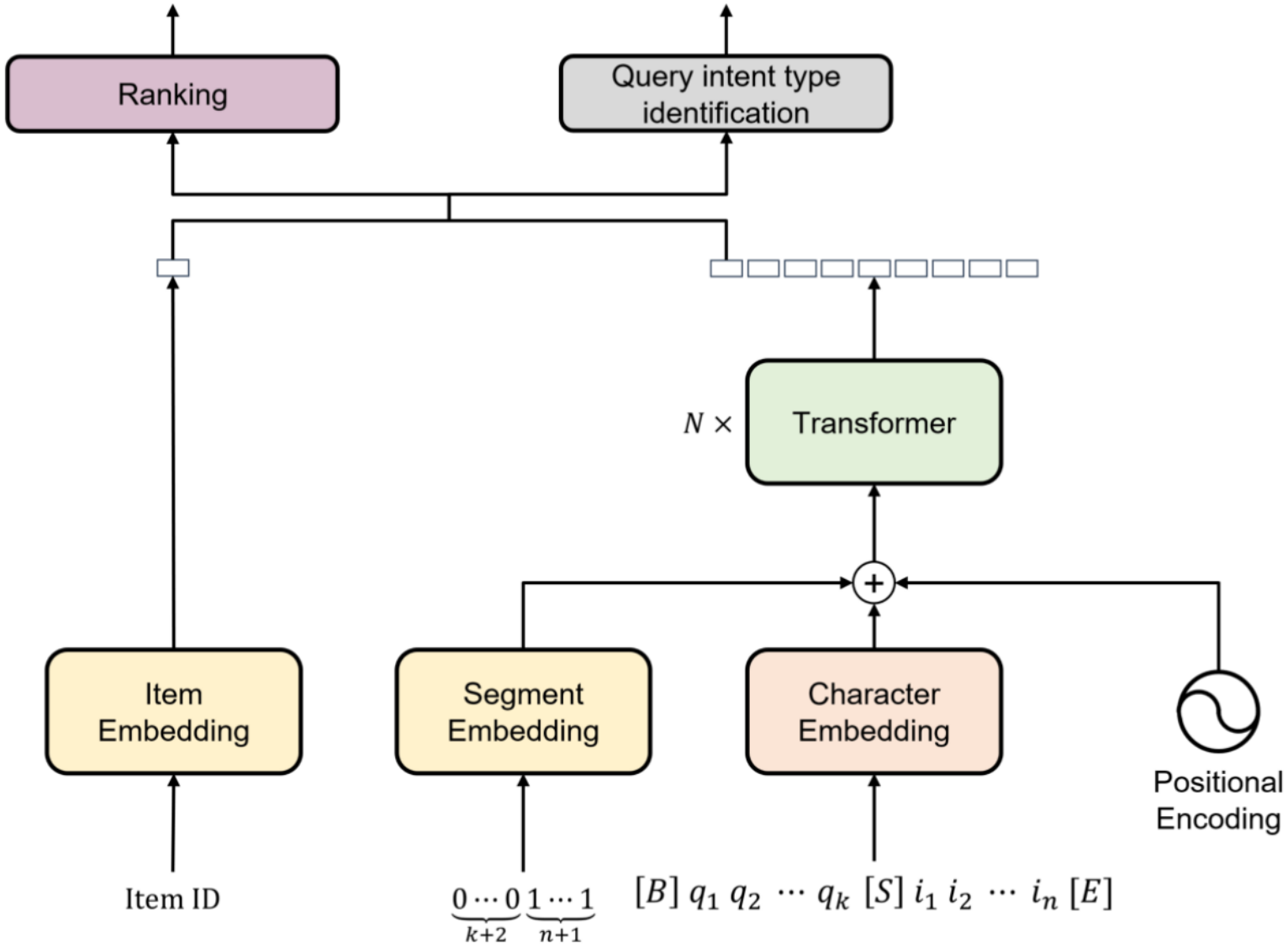
Log Analysis

Search Goal	Behavior	Rel. diff to music
Listen	Stream	+3.13%
Organize	Add to collection	+29.57%
	Add to playlist	-59.12%
	Follow artist	-92.37%
	Follow playlist	-39.88%
	Download	+593.02%
Share	Share link	+44.44%

Effort type	Rel. diff to music
Avg. deletions (# characters)	+53.25%
Avg. query length (# words)	+0.38%
Avg. query length (# characters)	+12.63%

Architecture

- Input
  - item ID
  - query
  - item title
- Output
  - ranking
  - query type



Method

Problem Formulation

$$D = \{(q_1, I_1, T_1, R_1), (q_2, I_2, T_2, R_2), ..., (q_N, I_N, T_N, R_N)\}$$

- q_i: the query text for the i^{th} query
- $I_i = \{I_{i1}, I_{i2}, ..., I_{in}\}$: the set of n candidates items for the query q_i
- T_i: the last interacted item type for the query q_i
- the j^{th} element in R_i denotes the relevance label for the item I_{ij}

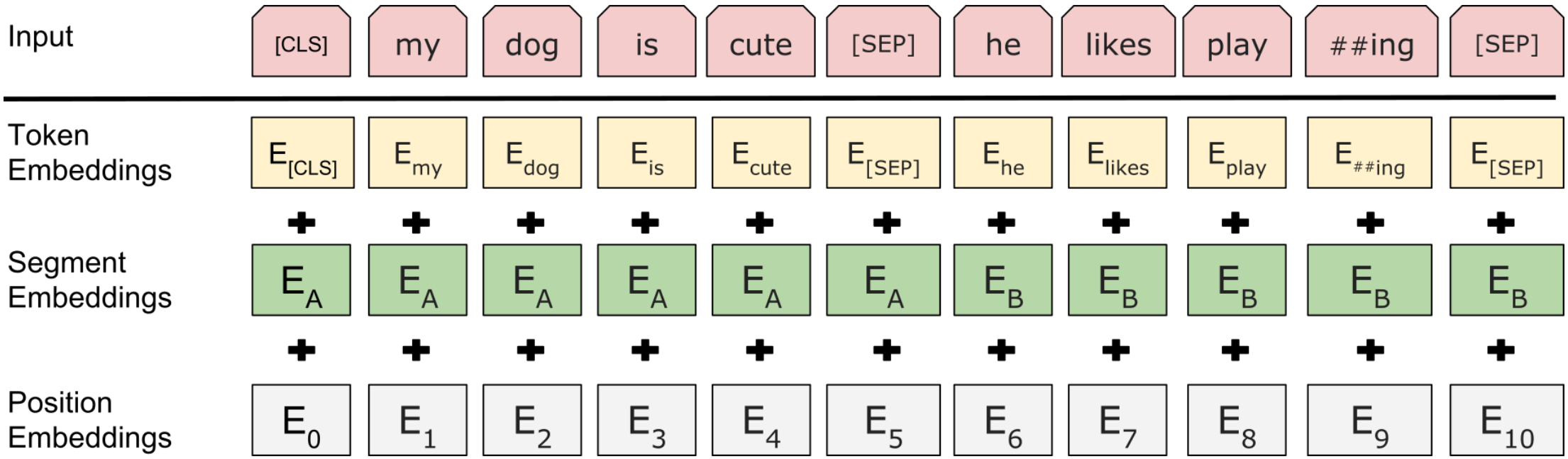
Segment Embedding

- first sequence(query):0
- second sequence(item title):1

Positional Embeddings

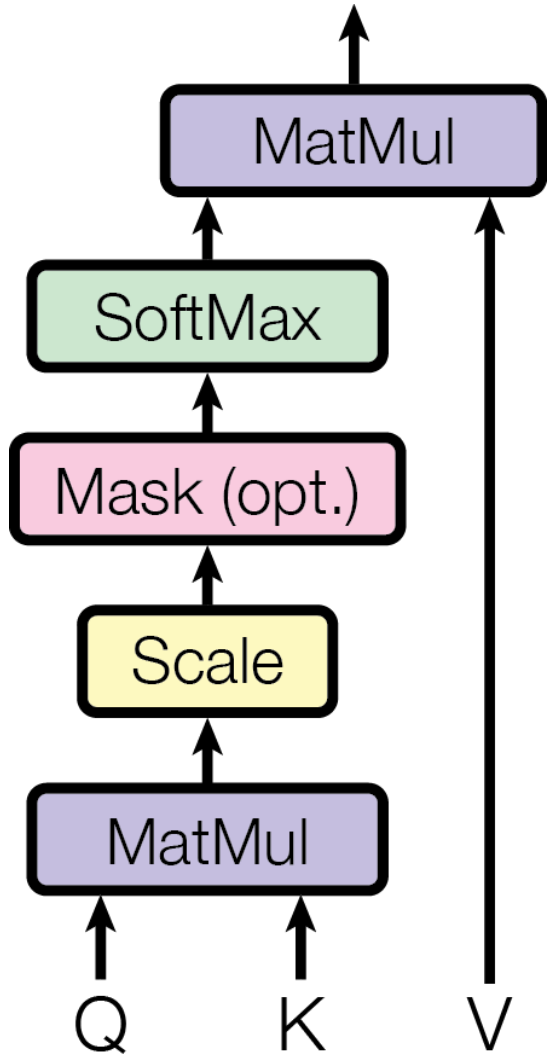
- $\mathrm{PE}_{(\mathrm{pos},2i)} = \sin(\frac{\mathrm{pos}}{10000^{2i/d_{\mathrm{model}}}})$
- $\mathrm{PE}_{(\mathrm{pos},2i+1)} = \cos(\frac{\mathrm{pos}}{10000^{2i/d_{\mathrm{model}}}})$

Input of Transformer



Scaled Dot-Product Attention

- $Z = \mathrm{softmax}(\frac{Q \times K^T}{\sqrt{d}})V$
- d is the embedding dimensionality



Loss Function

$$L_{\text{ranking}} = -R_{ij} \log \hat{R_{ij}} - (1 - R_{ij}) \log (1 - \hat{R_{ij}})$$

$$L_{\text{intent type}} = -T_i \log \hat{T_i} - (1 - T_i) \log (1 - \hat{T_i})$$

$$L = L_{\text{ranking}} + L_{\text{intent type}}$$

Experiment

Dataset

- sample logs from sessions over the period of one week in June 2020
- 100M sessions
- at least one click on music or podcasts

Training	Test	Validation
80%	10%	10%
6M query prefixes	**750K** query prefixes	**750K** query prefixes

Dataset

session id	query prefix	rank	candidate	click	query intent type
1	d	1	**D**rake	0	music
1	dr	2	**Dr**ake	1	music
2	b	1	**B**eyonce	0	music
3	d	1	**D**rake	0	music
3	da	2	**Da**vid Bowie	0	music
3	dai	3	**Dai**sies	0	music
3	dail	4	**Dail**y show	1	podcast

Evaluation Metrics

- MRR
- NDCG
- MAP

Baselines for Comparison

- Prefix Match and Item Popularity
- BERT Re-Ranking Model
- Character-based LSTM

Comparison Against Baselines

Model	NDCG@10	RPrec	MRR	MAP
PMIP	0.6264	0.3216	0.5573	0.5353
BERT	0.6468	0.3728	0.5852	0.5711
LSTM-Char	0.6522	0.3662	0.5886	0.5673
NIS	**0.6740***	**0.4056***	**0.6245***	**0.5968***

Comparison Against Baselines

Model	NDCG@10	RPrec	MRR	MAP
PMIP	0.6264	0.3216	0.5573	0.5353
BERT	0.6468	0.3728	0.5852	0.5711
LSTM-Char	0.6522	0.3662	0.5886	0.5673
NIS	**0.6740***	**0.4056***	**0.6245***	**0.5968***

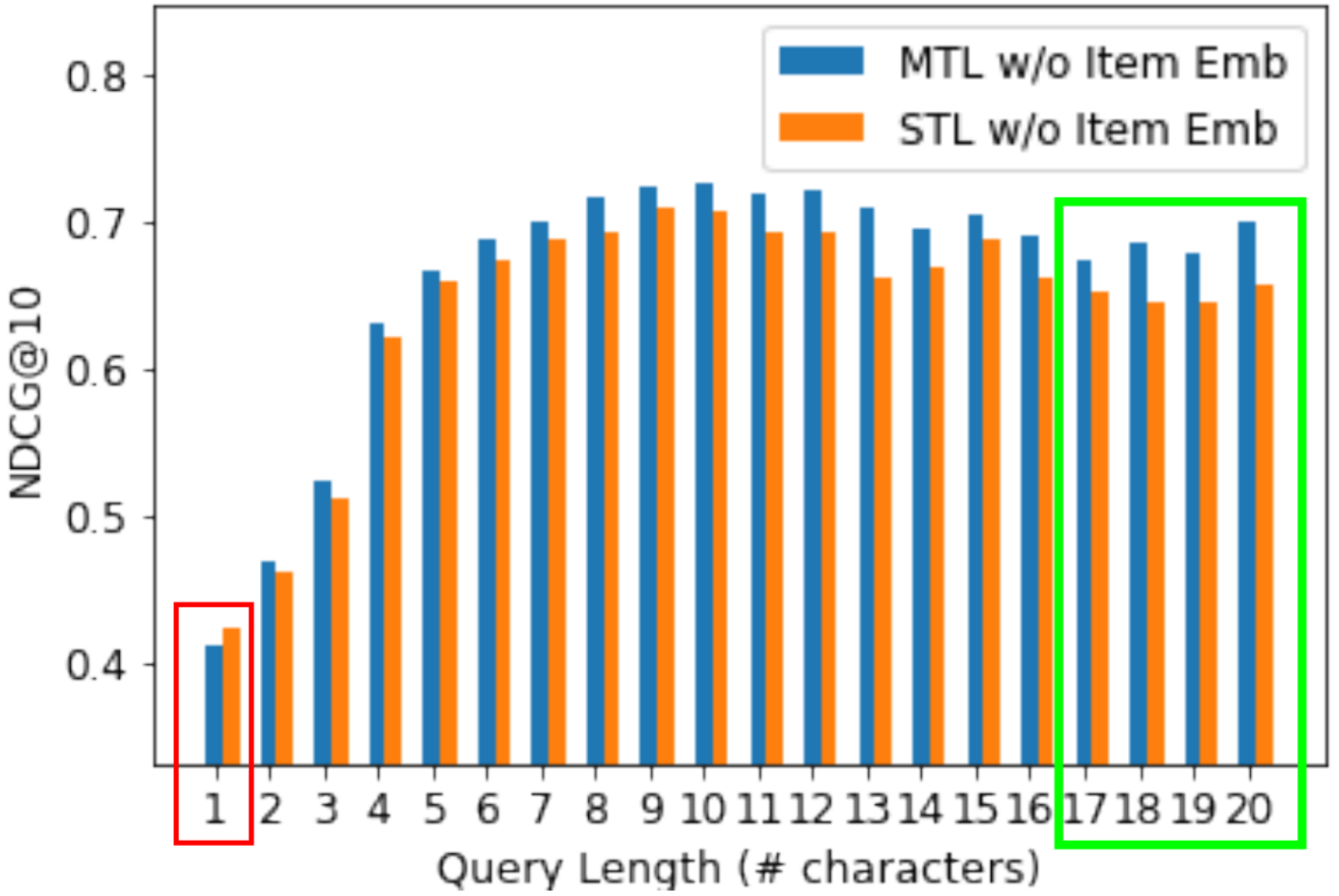
Ablation Study

ID	Model	NDCG@10	RPrec	MRR	MAP
–	NIS	0.6740^{123}	0.4056^{123}	0.6245^{123}	0.5968^{123}
1	NIS-query intent type identification (STL)	0.6630^{3}	0.3902^{3}	0.6013^{3}	0.5852^{3}
2	NIS-item embedding	0.6618^{3}	0.3910^{3}	0.5994^{3}	0.5849^{3}
3	NIS- query intent type identification - item embedding	0.6581	0.3823	0.5954	0.5791

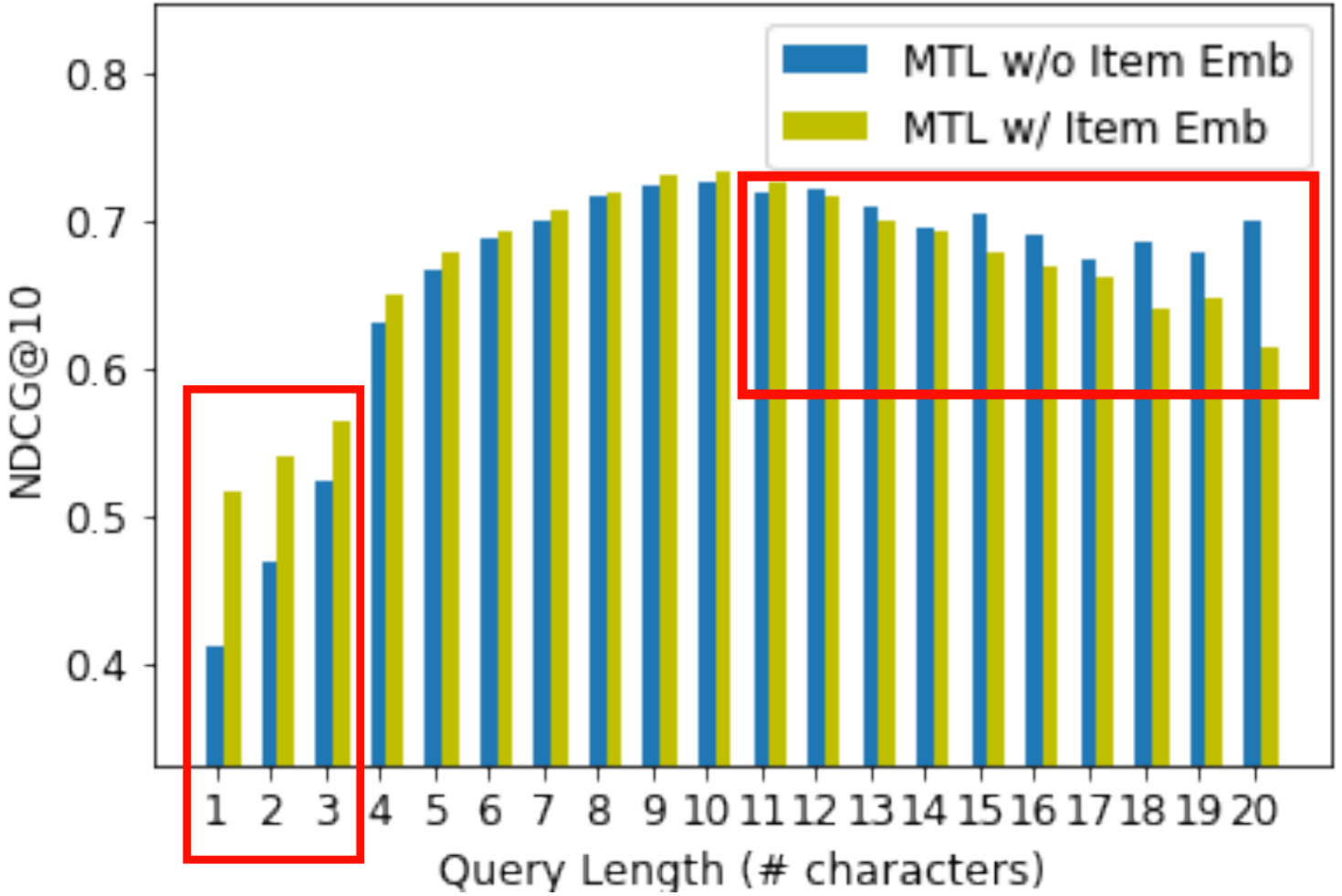
Performance by Query Intent Types

Model	Music Queries				Podcast Queries			
	NDCG@10	RPrec	MRR	MAP	NDCG@10	RPrec	MRR	MAP
LSTM-Char	0.6586	0.3684	0.6120	0.5736	0.6325	0.3633	0.5169	0.5425
NIS	0.6801*†	0.4087*†	0.6360*†	0.6034*†	0.6618*†	0.4098*	0.5559*	0.5787*†
NIS- item embedding	0.6675	0.3926	0.6087	0.5907	0.6557	0.4082	0.5527	0.5737
NIS- query intent type identification - item embedding	0.6533	0.3589	0.6021	0.5657	0.6323	0.3504	0.5695	0.5157

Performance by Query Length



Performance by Query Length



Conclusion

- Proposed a instant search model that matches query prefixes to any part of the items using multi-task learning
- The model outperforms strong baselines for both music and podcasts queries
- The multi-task learning and the item embedding component in the model contribute to ranking metrics individually, and also complement each other for improving model performance

References

- https://machinelearningmastery.com/a-gentle-introduction-to-positional-encoding-in-transformer-models-part-1/
- https://kazemnejad.com/blog/transformer_architecture_positional_encoding/
- https://research.atspotify.com/neural-instant-search-for-music-and-podcasts/