



Università degli Studi di Bari Aldo Moro

Dataset LFM-1b_artist

Emanuele Fontana

Tirocinio tesi triennale in Informatica
Anno accademico 2023/2024

LFM1b_artist è un dataset contenente informazioni riguardanti le interazioni tra utenti e artisti musicali.

DETASET ORIGINALE

STATISTICHE DATASET

Descrizione del dataset

Feature	Descrizione
n_users	120322
n_items	3123496
n_inter	65133026
sparsity	0.9998266933373666

Tabella 1: Informazioni sul dataset LFM1b_artist

Descrizione del knowledge graph

n_ent_head	823213
n_ent_tail	353607
n_rel	8
n_triple	2114049

Tabella 2: Informazioni sul knowledge graph del dataset LFM1b_artist

I nomi delle relazioni presenti nel knowledge graph sono i seguenti:

- music.recording.artist
- music.recording.releases
- music.recording.producer
- music.recording.engineer
- music.recording.featured_artists
- music.featured_artist.recordings
- music.release.artist
- music.artist.release

DATASET PROCESSATO

DESCRIZIONE

Il dataset originale risultava essere troppo grande per le risorse a nostra disposizione, dunque è stato opportunamente processato. In particolare sono state svolte le seguenti operazioni

- **Filtraggio:** il dataset è stato filtrato eliminando tutte le interazioni in cui erano coinvolti utenti e/o item con meno di 5 interazioni
- **Sampling:** dopo la fase di filtraggio, è stato effettuato un sampling casuale il cui scopo era quello di ridurre il numero di interazioni. In particolare sono stati selezionati casualmente 20000 utenti e 50000 item e sono state mantenute solo le interazioni in cui erano coinvolti utenti e gli item selezionati

STATISTICHE DATASET

Descrizione del dataset

Feature	Descrizione
n_users	19481
n_items	42547
n_inter	900212
sparsity	0.0.9989313587429705

Tabella 3: Informazioni sul dataset LFM1b_artist processato

Descrizione del knowledge graph

n_ent_head	15509
n_ent_tail	35156
n_rel	5
n_triple	46827

Tabella 4: Informazioni sul knowledge graph del dataset LFM1b_artist processato

I nomi delle relazioni presenti nel knowledge graph sono i seguenti:

-
- music.recording.artist
 - music.recording.releases
 - music.recording.producer
 - music.recording.engineer
 - music.recording.featured_artists

ESECUZIONE DEI MODELLI E RISULTATI

I modelli di recommendation sono stati eseguiti utilizzando questi parametri:

- **Numero di epoche:** 1
- **Train batch size:** 4096
- **Eval batch size:** 1024
- **Numero neighbors per l'ItemKNN:** 20

Qui di seguito sono riportati i risultati ottenuti dai modelli di recommendation.

Nome modello	Esito	Motivo fallimento
ItemKNN	Successo	
Pop	Successo	
Random	Fallito	Expected all tensors to be on the same device, but found at least two devices, cuda:0 and cpu! (when checking argument for argument tensors in method wrapper_CUDA_cat)
Simplex	Successo	
ADMMSLIM	Fallito	Terminato in modo anomalo
BPR	Successo	
DMF	Successo	
ENMF	Successo	
FISM	Successo	
NCEPLRec	Successo	
SLIMElasticNet	Fallito	Expected all tensors to be on the same device, but found at least two devices, cuda:0 and cpu! (when checking argument for argument tensors in method wrapper_CUDA_cat)
CDAE	Successo	
CONvNCF	Successo	
DiffRec	Successo	
EASE	Fallito	Terminato in modo anomalo
GCMC	Successo	
LDiffRec	Successo	
MacridVAE	Fallito	CUDA out of memory. Tried to allocate 664.00 MiB
MultiDAE	Successo	
MultiVAE	Successo	
NAIS	Fallito	CUDA out of memory. Tried to allocate 3.97 GiB
NeuMF	Successo	
NGCF	Successo	
NNCF	Successo	
LightGCN	Successo	
RecVAE	Rimandato	Tempo di esecuzione non accettabile
DGCF	Successo	
LINE	Successo	
NCL	Fallito	No module faiss
SGL	Successo	
SpectralCF	Successo	
CKE	Successo	
CFKG	Successo	
KGIN	Fallito	mismatch di versione tra PyTorch e torchscatter
KGNNLS	Successo	
KTUP	Successo	
MKR	Successo	
RippleNet	Successo	

Tabella 5: Esiti di esecuzione dei modelli