



UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

SWAP
researchgroup

Tesi in Metodi per il Ritrovamento dell'informazione

**ANALISI DELLA SOSTENIBILITÀ DEI SISTEMI DI RACCOMANDAZIONE A STATO
DELL'ARTE E PROPOSTA DI UN ADDESTRAMENTO SOSTENIBILE**

Relatore: Prof. Pasquale Lops


Relatore: Prof. Cataldo Musto

Correlatore: Dott. Giuseppe Spillo




Laureando: Emanuele Fontana

Università degli Studi di Bari Aldo Moro

Sostenibilità e AI

-  **Sostenibilità:** Capacità di soddisfare i bisogni delle generazioni presenti senza compromettere le generazioni future. Esempio di impegno è l'Agenda 2030 dell'ONU.



-  **Sostenibilità ambientale:** Uno degli aspetti della sostenibilità. Riduzione delle emissioni di CO2, rispetto delle risorse naturali e dell'ambiente.
-  **Green AI:** Durante l'addestramento si considera l'impatto ambientale in termini di emissioni di CO2 e consumo di energia.
-  **Red AI:** non considerava l'impatto ambientale (massime prestazioni)





Recommender Systems - Introduzione

- Software che suggerisce all'utente elementi di interesse basandosi sulle preferenze e i comportamenti passati.
- Basati su Intelligenza Artificiale.



Alcuni famose piattaforme che utilizzano sistemi di raccomandazione

Recommender Systems - Tipologie

-  **Collaborative Filtering**: basato sulle preferenze degli utenti
-  **Content-based Filtering**: basato sulla descrizione del contenuto degli item.
-  **Knowledge-aware**: utilizzano conoscenza esterna (es. knowledge graph)
-  **Hybrid**: combinazione delle precedenti.

Recommender Systems - Sostenibilità

Ad oggi i sistemi di raccomandazione non considerano l'impatto ambientale in fase di addestramento, sono dunque modelli **Red AI** che puntano alle massime prestazioni senza considerare l'impatto ambientale in termini di CO₂.

DOMANDA

E' possibile migliorare la sostenibilità di un sistema di raccomandazione migliorando il trade-off tra prestazioni e emissioni di CO₂?

Domande di ricerca e lavoro svolto

- **RQ1:** Qual è il trade-off tra emissioni e performance dei modelli di raccomandazione a stato dell'arte¹?
- **RQ2:** E' possibile usare un criterio di early-stopping basato sulle emissioni per migliorare il trade-off tra emissioni e performance dei modelli di raccomandazione a stato dell'arte?
- **RQ3:** Quali parametri possono essere utilizzati in questi criteri per migliorare il trade-off?

¹Modelli classici a cui fare riferimento

CodeCarbon

Per il tracking delle emissioni è stata usata la libreria Python **CodeCarbon**, la quale usa l'equivalente di anidride carbonica (CO_2eq) per misurare le emissioni mediante la seguente formula.

$$\text{emission} = CI \cdot PC \quad (1)$$

dove CI è il Carbon Intensity e PC è il Power Consumption (cioè l'energia consumata). I valori di CI dipendono dalle diverse fonti di energia utilizzate durante la computazione (es. energia solare, energia eolica, etc.). Se s è la fonte di energia, e_s sono le emissioni per KW/h di energia e p_s è la percentuale di energia prodotta dalla fonte s , allora il CI è dato da:

$$CI = \sum_{s \in S} e_s \cdot p_s \quad (2)$$

Dataset e modelli utilizzati

Dataset




Nome	Utenti	Item	Preferenze
 MovieLens10M	69,878	10,677 film	10,000,054
 MovieLens1M	6,040	3,706 film	1,000,209
 LastFM	120,322	3,123,496 canzoni	65,133,026





Tabella: Descrizione dataset

*I dataset sono stati ridotti selezionando un numero limitato di utenti e di item a causa delle limitazioni dell'infrastruttura utilizzata.

Modelli

- **Modelli di raccomandazione Collaborative Filtering:** BPR, DMF, LINE, MultiDAE, LightGCN, ItemKNN, NCF, DGCF
- **Modelli di raccomandazione Knowledge-Aware:** CKE, KGCN, KGNNLS, CFKG

Come sono stati valutati i modelli?

-  **Recall**: capacità di raccomandare item rilevanti
-  **NDCG**: considera l'ordine degli item raccomandati
-  **Average Popularity**: misura quanto sono popolari in media gli item raccomandati
-  **Gini Index**: misura la distribuzione degli item raccomandati

RQ1 Benchmarking - Emissioni

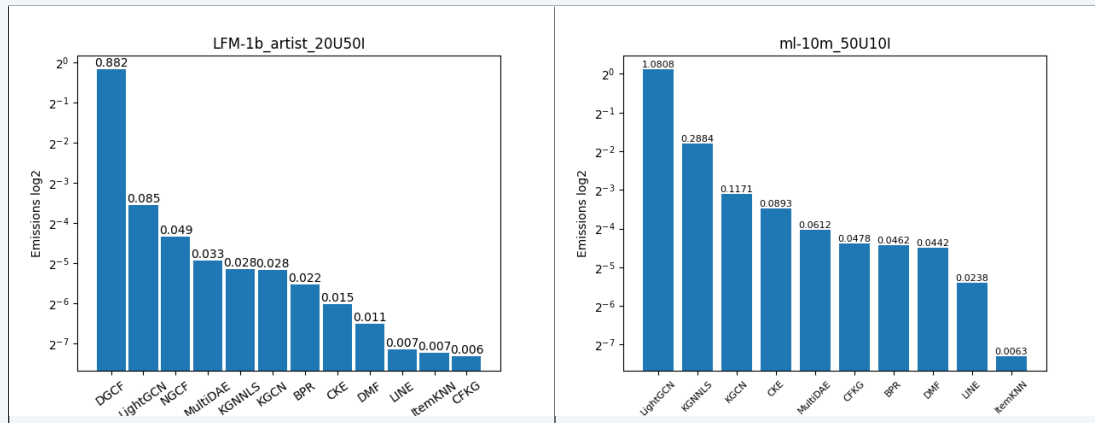
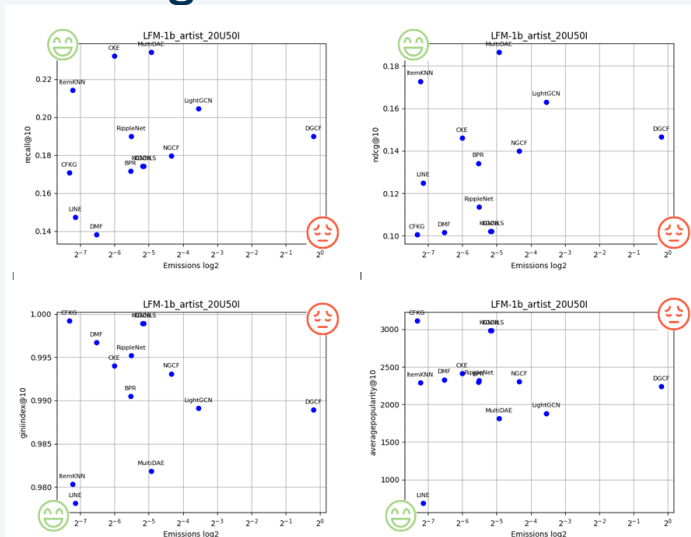


Tabella: Emissioni di CO2

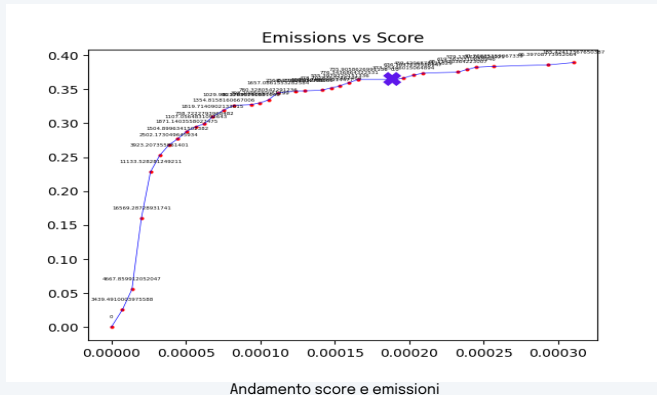
RQ1 Benchmarking - Trade-off



RQ2 Addestramento sostenibile - Introduzione

Sull'asse delle x troviamo le emissioni, sull'asse delle y lo score. Quando la derivata è al di sotto di una certa soglia **S** per un certo numero di epoche consecutive **E** l'addestramento termina (comportamento asintotico). Approssimazione della derivata della curva:

$$\frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}$$



RQ2 Addestramento sostenibile - Esplorazione

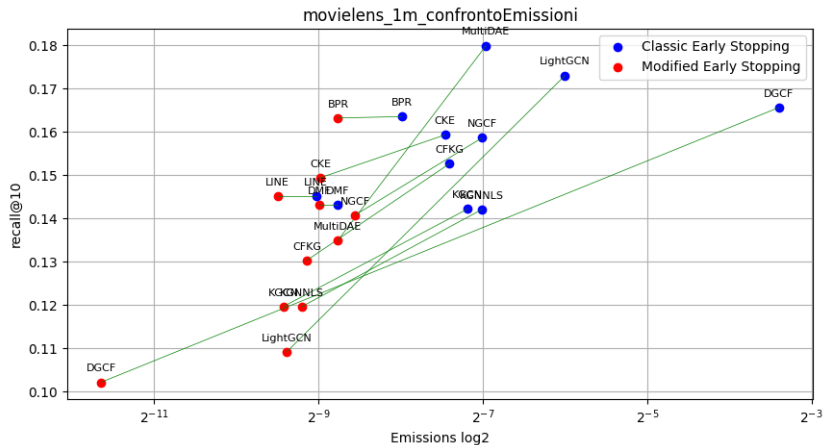
Modello	Emissioni criterio classico (g)	Emissioni criterio nuovo (g)	Riduzione	% riduzione emissioni
DMF	2.2927	1.97	0.32	14.21
LINE	1.91	1.38	0.53	27.69
NGCF	7.76	2.66	5.09	65.71
DGCF	94.90	10.46	84.44	88.97

Tabella: Esempi di confronto emissioni

Metrica	Modello	Score criterio classico	Score criterio nuovo	Riduzione	% riduzione score
recall@10	DMF	0.14	0.14	0.0	0.0
	LINE	0.15	0.15	0.0	0.0
	NGCF	0.15	0.14	0.01	6.67
	DGCF	0.17	0.10	0.07	41.18

Tabella: Esempi di confronto score

RQ2 Addestramento sostenibile - Esplorazione



Esempi di decrementi riscontrati sulla recall@10

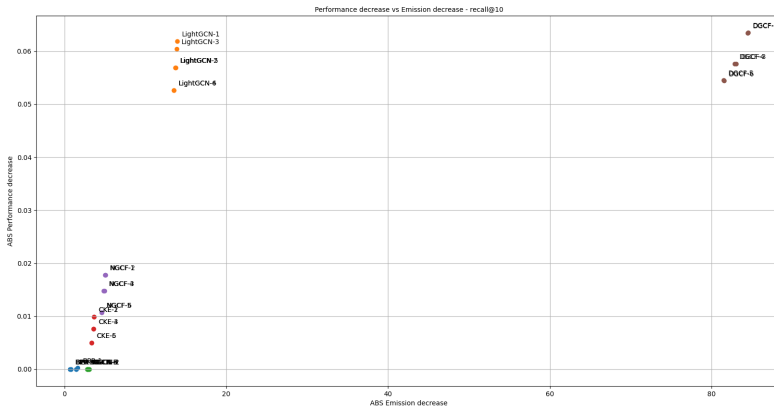
RQ3 Addestramento sostenibile - Confronto criteri

Sono stati eseguiti diversi esperimenti sul dataset MovieLens1M per confrontare diverse configurazioni dei parametri di soglia e di epoche, di seguito alcuni esempi

Modello	(Soglia,Epoche)	Riduzione emissioni	Riduzione score recall@10
DMF	(40,5)	0.30	0
	(30,7)	0.31	0
LINE	(40,5)	0.48	0
	(30,7)	0.47	0
NGCF	(40,5)	5.0	0.02
	(30,7)	4.63	0.01
DGCF	(40,5)	84.48	0.06
	(30,7)	81.52	0.05

Tabella: Esempi di risultati ottenuti

RQ3 Addestramento sostenibile - Confronto criteri



RQ3 Addestramento sostenibile - Risultati confronto criteri

Modello	Parametro più impattante	Migliori risultati
BPR	Soglia	Soglia 40 e 6 epoche
CFKG	Soglia	Soglia 40 e 6 epoche
CKE	Epoche consecutive	Soglia 40 e 6 epoche
DMF	Nessuno predominante	Soglia 40 e 7 epoche
KGCN	Epoche consecutive	Soglia 40 e 5 epoche
KGNNLS	Soglia	Soglia 40 e 5 epoche
LINE	Soglia	Soglia 40 e 7 epoche
MultiDAE	Soglia	Soglia 40 e 7 epoche
LightGCN	Soglia	Soglia 40 e 6 epoche
NGCF	Epoche consecutive	Soglia 40 e 5 epoche
DGCF	Epoche consecutive	Soglia 40 e 6 epoche

Tabella: Parametri più impattanti e migliori risultati per ciascun modello

RQ3 Addestramento sostenibile - Risultati confronto criteri

Tipo di Modello	Parametro predominante	Numero di Modelli	Modelli
Collaborative Filtering	Soglia	5	BPR, DMF, LightGCN, MultiDAE, LINE
Collaborative Filtering	Epoche	2	NGCF, DGCF
Knowledge-Aware	Soglia	2	CFKG, KGNNLS
Knowledge-Aware	Epoche	2	CKE, KGCN

Tabella: Riassunto dei parametri dominanti per tipo di modello

Conclusioni

Benchmarking

Si dimostra come spesso i modelli più complessi hanno emissioni maggiori non giustificate da un miglioramento delle performance elevato.

Addestramento sostenibile

E' possibile ridurre le emissioni di un modello di raccomandazione senza perdere in modo significativo di performance

Sviluppi futuri

Benchmarking

E' necessario effettuare più esperimenti variando dataset, modelli e hardware per avere una visione più completa del problema.

Addestramento sostenibile

Eseguire più esperimenti con altri dataset e altri hardware per confermare o meno i risultati ottenuti.

Iperparametri

Tutti gli esperimenti sono stati effettuati con iperparametri di default. Dunque tutta la fase di benchmarking e di addestramento sostenibile potrebbe essere rivista anche in termini di ricerca degli iperparametri migliori.

Grazie per l'attenzione!

Relatore: Prof. Pasquale Lops

Relatore: Prof. Cataldo Musto

Correlatore: Dott. Giuseppe Spillo

Laureando: Emanuele Fontana

Università degli Studi di Bari Aldo Moro