The 1st International Workshop on Deep Multimodal Learning for Information Retrieval

Ottawa, ON, Canada, 11-02-2023

Co-located with ACM Multimedia 2023

On Popularity Bias of Multimodal-aware Recommender Systems: a Modalities-driven Analysis

Daniele Malitesta, Giandomenico Cornacchia, Claudio Pomo, Tommaso Di Noia

Politecnico di Bari, Bari (Italy)

email: firstname.lastname@poliba.it

Politecnico di Bari

SISINF LAB

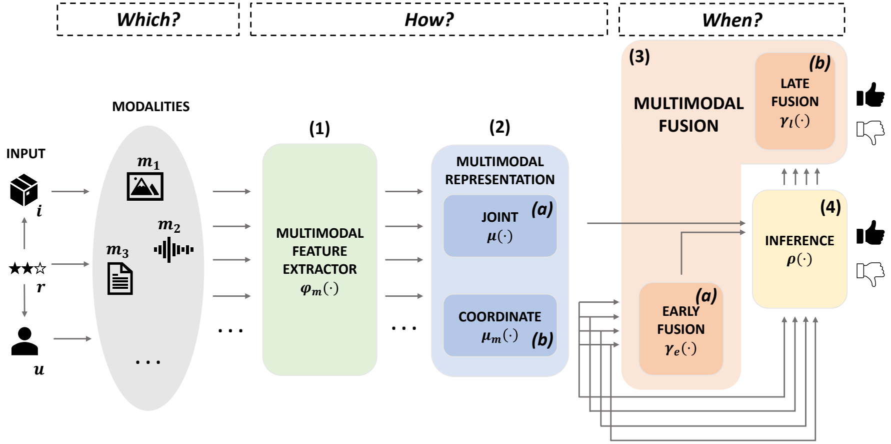
Outline

- Introduction and motivations
- Background
- Proposed analysis
- Results and discussion
- Conclusion and future work

Introduction and motivations

Recommendation systems leveraging multimodal data

Multimodal-aware **recommender systems** [Malitesta et al. (2023a)] exploit **multimodal** (i.e., audio, visual, textual) content **data** to augment the **representation** of **items**, thus **tackling** known **issues** such as dataset **sparsity** and the **inexplicable nature** of users' **actions** (i.e., views, clicks) on online **platforms**.

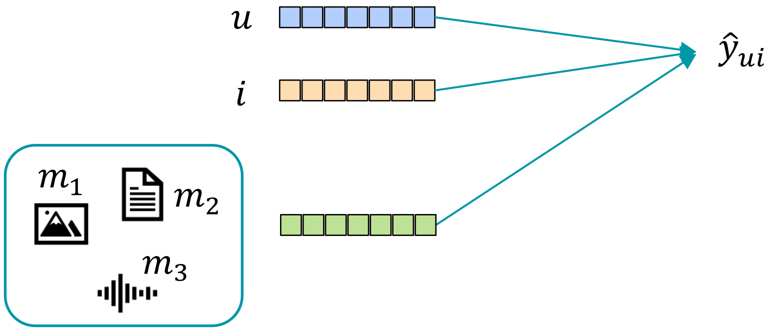


[Malitesta et al. (2023a)] 2023. *Formalizing Multimedia Recommendation through Multimodal Deep Learning*. Under review at TORS. Available online at: arXiv:2309.05273.

Multimodal-aware recommendation and factorization models

Most of **multimodal**-aware **recommender systems** are **based** upon **factorization models** for recommendation, such as the **matrix factorization** with **Bayesian personalized ranking** architecture (**MFBPR** **[Rendle et al.]**).

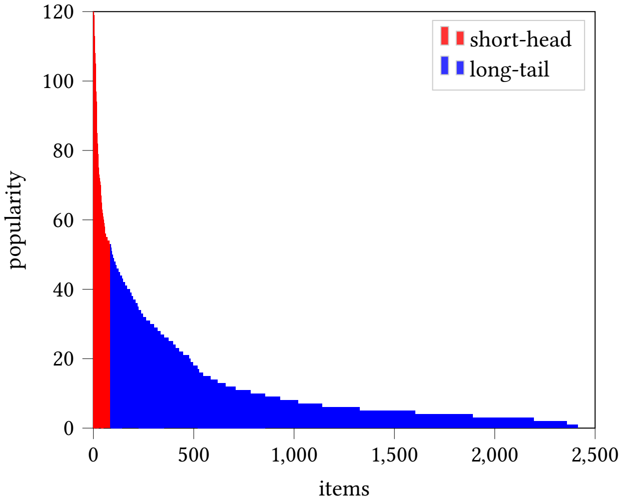
Given its **simple implementation** and **efficacy**, **MFBPR** has long constituted the **backbone** of **recommendation algorithms** in **collaborative filtering** **[He et al. (2020), Mao et al.]**, not only in multimodal recommendation.



[Rendle et al.] 2009. *BPR: Bayesian Personalized Ranking from Implicit Feedback.* In UAI.
[He et al. (2020)] 2020. *LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation*. In SIGIR. ACM, 639–648.
[Mao et al.] 2021. *SimpleX: A Simple and Strong Baseline for Collaborative Filtering*. In CIKM. ACM, 1243–1252.

Popularity bias in matrix factorization

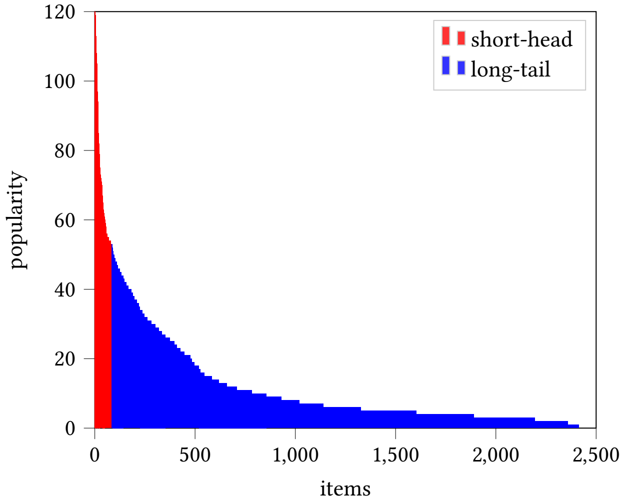
Nevertheless, the **literature** has shown that **MFBPR-like** models may be **affected** by **popularity bias** **[Abdollahpouri et al., Ricardo Baeza-Yates, Boratto et al., Jannach et al.]**. Such recommender systems tend to **boost** the **performance** of items from the *short-head* at the **detriment** of the items from the *long-tail*.



[Abdollahpouri et al.] 2017. *Controlling Popularity Bias in Learning-to-Rank Recommendation*. In RecSys. ACM, 42–46.
[Ricardo Baeza-Yates] 2020. *Bias in Search and Recommender Systems*. In RecSys. ACM, 2.
[Boratto et al.] 2021. *Connecting user and item perspectives in popularity debiasing for collaborative recommendation*. Inf. Process. Manag. 58, 1 (2021), 102387.
[Jannach et al.] 2015. *What recommenders recommend: an analysis of recommendation biases and possible countermeasures*. User Model. User Adapt. Interact. 25, 5 (2015), 427–491.

Popularity bias in multimodal-aware recommendation

Some recent works [Liu et al., Kowald and Lacic, Malitesta et al. (2023b)] **address bias** in **multimodal**-aware **recommendation**, but with **different definitions** and **settings** with respect to the one of **popularity bias** we presented **earlier**.



[Liu et al.] 2022. *EliMRec: Eliminating Single-modal Bias in Multimedia Recommendation*. In ACM Multimedia. ACM, 687–695.
[Kowald and Lacic] 2022. *Popularity Bias in Collaborative Filtering-Based Multimedia Recommender Systems.* In BIAS (Communications in Computer and Information Science, Vol. 1610). Springer, 1–11.
[Malitesta et al. (2023b)] 2023. *Disentangling the Performance Puzzle of Multimodal-aware Recommender Systems.* In EvalRS@KDD (CEUR Workshop Proceedings, Vol. 3450). CEUR-WS.org.

Our contributions

- ✓ **Propose** one of the **first analyses** on how **multimodal**-aware **recommender** systems may **amplify popularity bias**
- ✓ Select **four state-of-the-art multimodal**-aware **recommender** systems (i.e., VBPR, MMGCN, GRCN, and LATTICE)
- ✓ **Train** them on **three categories** of the **Amazon Catalogue** (i.e., Office, Toys, and Clothing)
- ✓ **Evaluate** the performance on **recommendation accuracy** and **popularity bias** (i.e., **diversity** and percentage of **retrieved items** from the **long-tail**)
- ✓ **Assess** the **separate impact** of each **multimodal side information** on **single** and **paired** recommendation **metrics**

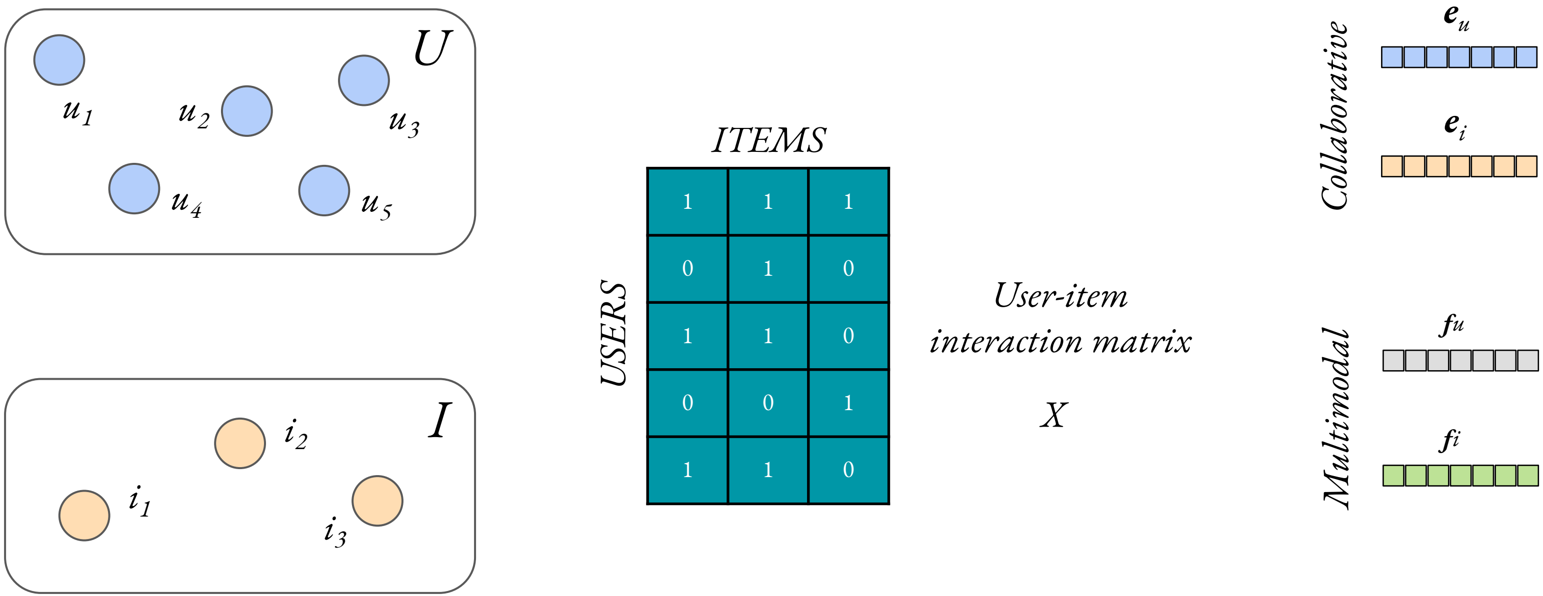
Research questions

RQ1) How do multimodal-aware recommendation models behave in terms of accuracy, diversity, and popularity bias?

RQ2) What is the influence of each modality (i.e., visual, textual, multimodal) on such performance measures?

Background

Preliminaries



Multimodal-aware recommender systems

- Visual Bayesian personalized ranking (VBPR **[He et al. (2016)]**)
- Multimodal graph convolutional network for recommendation (MMGCN **[Wei et al. (2019)]**)
- Graph-refined convolutional network (GRCN **[Wei et al. (2020)]**)
- Latent structure mining method for multimodal recommendation (LATTICE **[Zhang et al.]**)

Models	**Year**	**Venue**	**Prediction**
VBPR	2016	AAAI	$\hat{x}_{ui} = \mathbf{e}_u^\top \mathbf{e}_i + \mathbf{f}_u^\top t(\mathbf{f}_i) \quad \text{with} \quad \mathbf{f}_i = \Vert_{m \in \mathcal{M}} \mathbf{f}_i^m$
MMGCN	2019	MM	$\hat{x}_{ui} = \mathbf{f}_u^\top \mathbf{f}_i \quad \text{with} \quad \mathbf{f}_u = \sum_{m \in \mathcal{M}} c(\mathbf{e}_u, g(\mathbf{f}_u^m), t(\mathbf{f}_u^m, \mathbf{e}_u))$
GRCN	2020	MM	$\hat{x}_{ui} = \mathbf{f}_u^\top \mathbf{f}_i \quad \text{with} \quad \mathbf{f}_u = g(\mathbf{e}_u, \mathbf{f}_u^m, \forall m \in \mathcal{M}) \Vert \left(\Vert_{m \in \mathcal{M}} t(\mathbf{f}_u^m) \right)$
LATTICE	2021	MM	$\hat{x}_{ui} = \mathbf{e}_u^\top \mathbf{f}_i \quad \text{with} \quad \mathbf{f}_i = \mathbf{e}_i + \frac{g(\mathbf{e}_i, \mathbf{f}_i^m, \forall m \in \mathcal{M})}{\Vert g(\mathbf{e}_i, \mathbf{f}_i^m, \forall m \in \mathcal{M}) \Vert_2}$

[He et al. (2016)] 2016. *VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback.* In AAAI. AAAI Press, 144–150.
[Wei et al. (2019)] 2019. *MMGCN: Multi-modal Graph Convolution Network for Personalized Recommendation of Micro-video.* In ACM Multimedia. ACM, 1437–1445.
[Wei et al. (2020)] 2020. *Graph-Refined Convolutional Network for Multimedia Recommendation with Implicit Feedback.* In ACM Multimedia. ACM, 3541–3549.
[Zhang et al.] 2021. *Mining Latent Structures for Multimedia Recommendation*. In ACM Multimedia. ACM, 3872–3880.

Proposed analysis

Datasets and multimodal features

Amazon Catalogue [McAuley et al.]

Datasets	$\|\mathcal{U}\|$	$\|\mathcal{I}\|$	$\|\mathcal{R}\|$	Sparsity (%)
Office	4,905	2,420	53,258	99.5513
Toys	19,412	11,924	167,597	99.9276
Clothing	39,387	23,033	278,677	99.9693

Multimodal features

- Visual features: 4,096 embeddings **[Deldjoo et al.]**
- Textual features: 1,024 embeddings **[Zhang et al.]**

[McAuley et al.] 2015. *Image-Based Recommendations on Styles and Substitutes.* In SIGIR. ACM, 43–52.
[Deldjoo et al.] 2021. *A Study on the Relative Importance of Convolutional Neural Networks in Visually-Aware Recommender Systems*. In CVPR Workshops. Computer Vision Foundation / IEEE, 3961–3967.
[Zhang et al.] 2021. *Mining Latent Structures for Multimedia Recommendation*. In ACM Multimedia. ACM, 3872–3880.

Evaluation metrics

Accuracy

Recall: $\text{Recall@}k = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{|\text{Rel}_u\ @k|}{|\text{Rel}_u|},$

Normalized discount cumulative gain:

$$\text{nDCG@}k = \frac{1}{|\mathcal{U}|} \sum_{u} \frac{\text{DCG}_u@k}{\text{IDCG}_u@k},$$

Popularity Bias

Item coverage: $\text{iCov@}k = \frac{|\bigcup_u \hat{\mathcal{I}}_u@k|}{|\mathcal{I}_{train}|},$

Average percentage of long-tail items [Abdollahpouri et al.]

$$\text{APLT@}k = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{|\{i \mid i \in (\hat{\mathcal{I}}_u@k \cap \sim \Phi)\}|}{k},$$

[Abdollahpouri et al.] 2017. *Controlling Popularity Bias in Learning-to-Rank Recommendation*. In RecSys. ACM, 42–46.

Results and discussion

Recommendation accuracy, diversity, and popularity bias (RQ1)

Datasets	Models	top@10				top@20				top@50			
		Recall↑	nDCG↑	iCov↓	APLT↓	Recall↑	nDCG↑	iCov↓	APLT↓	Recall↑	nDCG↑	iCov↓	APLT↓
Office	Random	0.0034	0.0020	2,414	0.5950	0.0079	0.0034	2,414	0.5948	0.0220	0.0068	2,414	0.5924
	MostPop	0.0302	0.0208	20	0.0000	0.0533	0.0282	32	0.0000	0.1143	0.0439	66	0.0000
	MFBPR	0.0602	0.0389	2,268	0.2294	0.0955	0.0500	2,357	0.2379	0.1657	0.0677	2,398	0.2513
	VBPR	0.0652	0.0419	2,265	0.2321	0.1025	0.0533	2,354	0.2375	0.1774	0.0721	2,404	0.2469
	MMGCN	0.0455	0.0300	**74**	**0.0016**	0.0798	0.0405	**112**	**0.0078**	0.1575	0.0598	**247**	**0.0205**
	GRCN	0.0393	0.0253	2,390	0.3438	0.0667	0.0339	2,409	0.3469	0.1250	0.0488	2,414	0.3548
	LATTICE	**0.0664**	**0.0449**	2,121	0.1752	**0.1029**	**0.0566**	2,315	0.2039	**0.1780**	**0.0751**	2,397	0.2413
Toys	Random	0.0011	0.0006	11,879	0.4894	0.0021	0.0008	11,879	0.4896	0.0051	0.0015	11,879	0.4902
	MostPop	0.0130	0.0075	13	0.0000	0.0229	0.0104	24	0.0000	0.0451	0.0156	56	0.0000
	MFBPR	0.0641	0.0403	10,016	0.1167	0.0903	0.0481	10,944	0.1268	0.1394	0.0596	11,544	0.1460
	VBPR	0.0710	0.0458	10,085	0.1064	0.1006	0.0545	11,026	0.1180	0.1523	0.0667	11,624	0.1400
	MMGCN	0.0256	0.0150	**4,499**	0.0961	0.0426	0.0200	**6,238**	0.1058	0.0785	0.0285	**8,657**	0.1263
	GRCN	0.0554	0.0354	11,007	0.2368	0.0831	0.0436	11,609	0.2482	0.1355	0.0559	11,847	0.2679
	LATTICE	**0.0805**	**0.0512**	8,767	**0.0546**	**0.1165**	**0.0617**	10,285	**0.0684**	**0.1771**	**0.0759**	11,397	**0.0950**
Clothing	Random	0.0004	0.0002	23,016	0.4487	0.0010	0.0003	23,016	0.4478	0.0024	0.0006	23,016	0.4482
	MostPop	0.0089	0.0046	13	0.0000	0.0157	0.0063	24	0.0000	0.0322	0.0095	56	0.0000
	MFBPR	0.0303	0.0156	18,414	0.0729	0.0459	0.0195	20,582	0.0824	0.0734	0.0249	22,171	0.1017
	VBPR	0.0339	0.0181	19,195	0.0809	0.0529	0.0229	21,251	0.0915	0.0847	0.0292	22,555	0.1112
	MMGCN	0.0227	0.0119	**1,744**	**0.0044**	0.0348	0.0150	**2,864**	**0.0066**	0.0609	0.0201	**5,373**	**0.0121**
	GRCN	0.0319	0.0164	21,490	0.2358	0.0496	0.0209	22,503	0.2459	0.0858	0.0281	22,954	0.2631
	LATTICE	**0.0502**	**0.0275**	13,463	0.0134	**0.0744**	**0.0336**	17,538	0.0207	**0.1186**	**0.0425**	21,458	0.0385

LATTICE stands out for its accuracy performance...😃

...but amplifies popularity bias 😢

Recommendation accuracy, diversity, and popularity bias (RQ1)

Datasets	Models	top@10 Recall↑	top@10 nDCG↑	top@10 iCov↓	top@10 APLT↓	top@20 Recall↑	top@20 nDCG↑	top@20 iCov↓	top@20 APLT↓	top@50 Recall↑	top@50 nDCG↑	top@50 iCov↓	top@50 APLT↓
Office	Random	0.0034	0.0020	2,414	0.5950	0.0079	0.0034	2,414	0.5948	0.0220	0.0068	2,414	0.5924
	MostPop	0.0302	0.0208	20	0.0000	0.0533	0.0282	32	0.0000	0.1143	0.0439	66	0.0000
	MFBPR	0.0602	0.0389	2,268	0.2294	0.0955	0.0500	2,357	0.2379	0.1657	0.0677	2,398	0.2513
	VBPR	0.0652	0.0419	2,265	0.2321	0.1025	0.0533	2,354	0.2375	0.1774	0.0721	2,404	0.2469
	MMGCN	0.0455	0.0300	**74**	**0.0016**	0.0798	0.0405	**112**	**0.0078**	0.1575	0.0598	**247**	**0.0205**
	GRCN	0.0393	0.0253	2,390	0.3438	0.0667	0.0339	2,409	0.3469	0.1250	0.0488	2,414	0.3548
	LATTICE	**0.0664**	**0.0449**	2,121	0.1752	**0.1029**	**0.0566**	2,315	0.2039	**0.1780**	**0.0751**	2,397	0.2413
Toys	Random	0.0011	0.0006	11,879	0.4894	0.0021	0.0008	11,879	0.4896	0.0051	0.0015	11,879	0.4902
	MostPop	0.0130	0.0075	13	0.0000	0.0229	0.0104	24	0.0000	0.0451	0.0156	56	0.0000
	MFBPR	0.0641	0.0403	10,016	0.1167	0.0903	0.0481	10,944	0.1268	0.1394	0.0596	11,544	0.1460
	VBPR	0.0710	0.0458	10,085	0.1064	0.1006	0.0545	11,026	0.1180	0.1523	0.0667	11,624	0.1400
	MMGCN	0.0256	0.0150	**4,499**	0.0961	0.0426	0.0200	**6,238**	0.1058	0.0785	0.0285	**8,657**	0.1263
	GRCN	0.0554	0.0354	11,007	0.2368	0.0831	0.0436	11,609	0.2482	0.1355	0.0559	11,847	0.2679
	LATTICE	**0.0805**	**0.0512**	8,767	**0.0546**	**0.1165**	**0.0617**	10,285	**0.0684**	**0.1771**	**0.0759**	11,397	**0.0950**
Clothing	Random	0.0004	0.0002	23,016	0.4487	0.0010	0.0003	23,016	0.4478	0.0024	0.0006	23,016	0.4482
	MostPop	0.0089	0.0046	13	0.0000	0.0157	0.0063	24	0.0000	0.0322	0.0095	56	0.0000
	MFBPR	0.0303	0.0156	18,414	0.0729	0.0459	0.0195	20,582	0.0824	0.0734	0.0249	22,171	0.1017
	VBPR	0.0339	0.0181	19,195	0.0809	0.0529	0.0229	21,251	0.0915	0.0847	0.0292	22,555	0.1112
	MMGCN	0.0227	0.0119	**1,744**	**0.0044**	0.0348	0.0150	**2,864**	**0.0066**	0.0609	0.0201	**5,373**	**0.0121**
	GRCN	0.0319	0.0164	21,490	0.2358	0.0496	0.0209	22,503	0.2459	0.0858	0.0281	22,954	0.2631
	LATTICE	**0.0502**	**0.0275**	13,463	0.0134	**0.0744**	**0.0336**	17,538	0.0207	**0.1186**	**0.0425**	21,458	0.0385

MMGCN struggles with diversity... 🤒

...exhibits strong popularity bias... 😱

...and sacrifices accuracy in certain scenarios ☠️

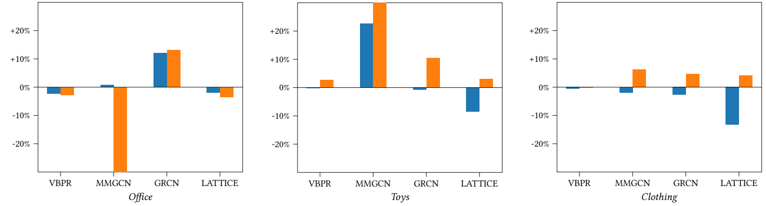
Recommendation accuracy, diversity, and popularity bias (RQ1)

Datasets	Models	top@10				top@20				top@50			
		Recall↑	**nDCG↑**	**iCov↓**	**APLT↓**	**Recall↑**	**nDCG↑**	**iCov↓**	**APLT↓**	**Recall↑**	**nDCG↑**	**iCov↓**	**APLT↓**
Office	Random	0.0034	0.0020	2,414	0.5950	0.0079	0.0034	2,414	0.5948	0.0220	0.0068	2,414	0.5924
	MostPop	0.0302	0.0208	20	0.0000	0.0533	0.0282	32	0.0000	0.1143	0.0439	66	0.0000
	MFBPR	0.0602	0.0389	2,268	0.2294	0.0955	0.0500	2,357	0.2379	0.1657	0.0677	2,398	0.2513
	VBPR	0.0652	0.0419	2,265	0.2321	0.1025	0.0533	2,354	0.2375	0.1774	0.0721	2,404	0.2469
	MMGCN	0.0455	0.0300	**74**	**0.0016**	0.0798	0.0405	**112**	**0.0078**	0.1575	0.0598	**247**	**0.0205**
	GRCN	0.0393	0.0253	2,390	0.3438	0.0667	0.0339	2,409	0.3469	0.1250	0.0488	2,414	0.3548
	LATTICE	**0.0664**	**0.0449**	2,121	0.1752	**0.1029**	**0.0566**	2,315	0.2039	**0.1780**	**0.0751**	2,397	0.2413
Toys	Random	0.0011	0.0006	11,879	0.4894	0.0021	0.0008	11,879	0.4896	0.0051	0.0015	11,879	0.4902
	MostPop	0.0130	0.0075	13	0.0000	0.0229	0.0104	24	0.0000	0.0451	0.0156	56	0.0000
	MFBPR	0.0641	0.0403	10,016	0.1167	0.0903	0.0481	10,944	0.1268	0.1394	0.0596	11,544	0.1460
	VBPR	0.0710	0.0458	10,085	0.1064	0.1006	0.0545	11,026	0.1180	0.1523	0.0667	11,624	0.1400
	MMGCN	0.0256	0.0150	**4,499**	0.0961	0.0426	0.0200	**6,238**	0.1058	0.0785	0.0285	**8,657**	0.1263
	GRCN	0.0554	0.0354	11,007	0.2368	0.0831	0.0436	11,609	0.2482	0.1355	0.0559	11,847	0.2679
	LATTICE	**0.0805**	**0.0512**	8,767	**0.0546**	**0.1165**	**0.0617**	10,285	**0.0684**	**0.1771**	**0.0759**	11,397	**0.0950**
Clothing	Random	0.0004	0.0002	23,016	0.4487	0.0010	0.0003	23,016	0.4478	0.0024	0.0006	23,016	0.4482
	MostPop	0.0089	0.0046	13	0.0000	0.0157	0.0063	24	0.0000	0.0322	0.0095	56	0.0000
	MFBPR	0.0303	0.0156	18,414	0.0729	0.0459	0.0195	20,582	0.0824	0.0734	0.0249	22,171	0.1017
	VBPR	0.0339	0.0181	19,195	0.0809	0.0529	0.0229	21,251	0.0915	0.0847	0.0292	22,555	0.1112
	MMGCN	0.0227	0.0119	**1,744**	**0.0044**	0.0348	0.0150	**2,864**	**0.0066**	0.0609	0.0201	**5,373**	**0.0121**
	GRCN	0.0319	0.0164	21,490	0.2358	0.0496	0.0209	22,503	0.2459	0.0858	0.0281	22,954	0.2631
	LATTICE	**0.0502**	**0.0275**	13,463	0.0134	**0.0744**	**0.0336**	17,538	0.0207	**0.1186**	**0.0425**	21,458	0.0385

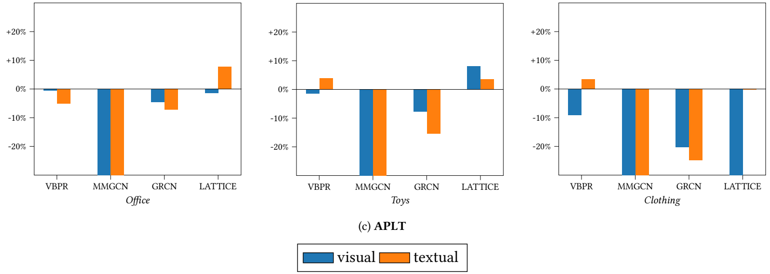
VBPR and GRCN better manage all the metrics by finding the right compromise among them 😎

Modalities influence on recommendation performance (RQ2)

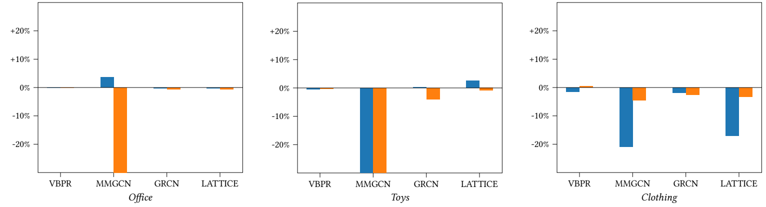
Single metric setting



(a) **Recall**



(c) **APLT**



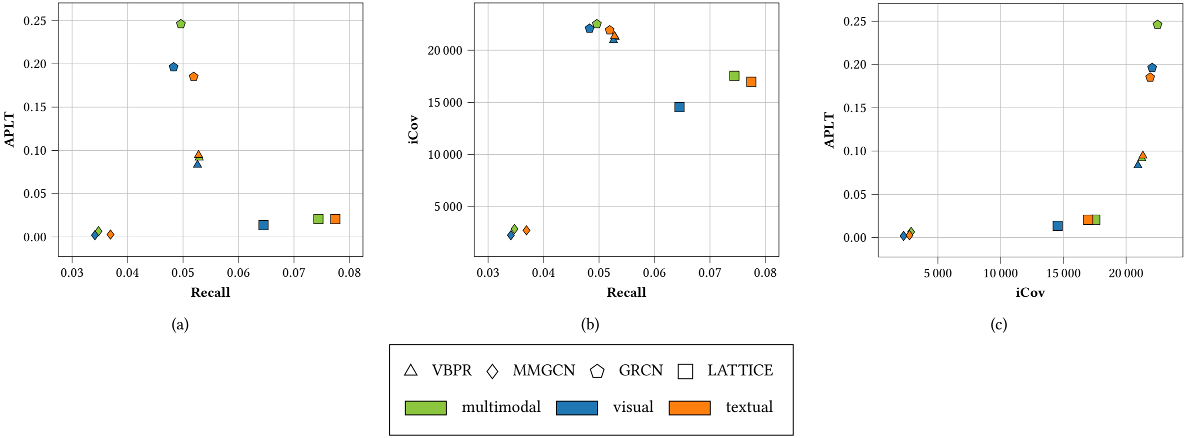
(b) **iCov**

The textual modality improves the accuracy… 💪

…while both modalities negatively affect the diversity and reinforce the popularity bias 😭

Modalities influence on recommendation performance (RQ2)

Pair-wise metric setting

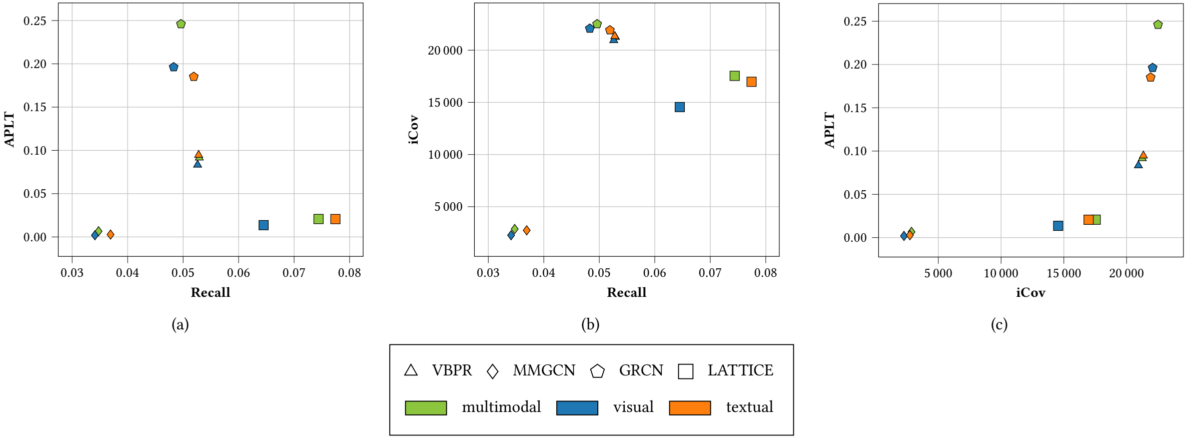


The textual modality has a significant influence on accuracy… 😣

but minimal effects on diversity and popularity bias 😇

Modalities influence on recommendation performance (RQ2)

Pair-wise metric setting (cont'd)



The visual modality reduces accuracy... 😨

...and jointly worsens the popularity bias and diversity 😵

Conclusion and future work

Conclusion

- Analysis on **influence** of **multimodality** on **popularity bias**
- **Four** SOTA **multimodal recommendation** approaches on **three datasets**
- Three **evaluation dimensions** and three **modality settings**
- **[RQ1] VBPR** and **GRCN** strike a **better compromise** among all metrics
- **[RQ2 single] Separate injection** of modalities **improves accuracy** but **negatively** impacts **diversity** and **popularity bias**
- **[RQ2 pairs textual] Highly impacts** on **accuracy** but **little effect** on **diversity** and **popularity bias**
- **[RQ2 pairs visual] Reduces accuracy** while **exacerbating popularity bias** and **limiting** the **diversity**

Future work

- **More complete study** on the performance of these models
- **Assessing** the **performance** of **more recent multimodal** approaches **[Malitesta et al. (2023a)]**

Reach us out!

The authors:

- **Daniele Malitesta** (daniele.malitesta@poliba.it)
- Giandomenico Cornacchia (giandomenico.cornacchia@poliba.it)
- Claudio Pomo (claudio.pomo@poliba.it)
- Tommaso Di Noia (tommaso.dinoia@poliba.it)

Don’t forget to check out our theoretical/experimental survey

Formalizing Multimedia Recommendation through Multimodal Deep Learning

DANIELE MALITESTA, Politecnico di Bari, Italy
GIANDOMENICO CORNACCHIA, Politecnico di Bari, Italy
CLAUDIO POMO, Politecnico di Bari, Italy
FELICE ANTONIO MERRA*, Amazon Science, Germany
TOMMASO DI NOIA, Politecnico di Bari, Italy
EUGENIO DI SCIASCIO, Politecnico di Bari, Italy

Recommender systems (RSs) provide customers with a personalized navigation experience within the vast catalogs of products and services offered on popular online platforms. Despite the substantial success of traditional RSs, recommendation remains a highly challenging task, especially in specific scenarios and domains. For example, human affinity for items described through multimedia content (e.g., images, audio, and text), such as fashion products, movies, and music, is multi-faceted and primarily driven by their diverse characteristics. Therefore, by leveraging all available signals in such scenarios, multimodality enables us to tap into richer information sources and construct more refined user/item profiles for recommendations. Despite the growing number of multimodal techniques proposed for multimedia recommendation, the existing literature lacks a shared and universal schema for modeling and solving the recommendation problem through the lens of multimodality. Given the recent advances in multimodal deep learning for other tasks and scenarios where precise theoretical and applicative procedures exist, we also consider it imperative to formalize a general multimodal schema for multimedia recommendation. In this work, we first provide a comprehensive literature review of multimodal approaches for multimedia recommendation from the last eight years. Second, we outline the theoretical foundations of a multimodal pipeline for multimedia recommendation by identifying and formally organizing recurring solutions/patterns. Third, we demonstrate its rationale by conceptually applying it to selected state-of-the-art approaches in multimedia recommendation.