Toward Effective Fashion Item Compatibility Modelling*

Extended Abstract

Matteo Attimonelli[1,†], Claudio Pomo[1,†], Dietmar Jannach[2] and Tommaso Di Noia[1]

[1]Politecnico di Bari, Italy
[2]University of Klagenfurt, Austria

Abstract

The rise in online fashion retail has led to increased research in fashion compatibility modeling and item retrieval. These technologies help users find fashion items based on text descriptions or reference images, focusing on how well items go together. However, retrieving complementary items is challenging due to the need for precise compatibility models. We propose the Compatibility-to-Retrieval Model (C2RM), which aims to improve fashion image retrieval using image-to-image translation. First, a Conditional Generative Adversarial Network generates target items from query items. Next, these generated samples are fed into C2RM, enhancing compatibility modeling and retrieval accuracy exploiting the first and second step. Evaluations on two datasets show C2RM's superior performance over current baselines.

Keywords

Fashion Recommendation, Compatibility Modeling, Generative Models

1. Introduction

In e-commerce, particularly within the fashion industry, there is a growing emphasis on personalized and retrieval models due to the rising demand for custom recommendations [1], particularly those employing a multimodal approach [2]. Image retrieval systems, designed to fetch items based on user queries through textual descriptions or reference images, have advanced significantly. Complementary item retrieval [3, 4, 5], such as top-bottom retrieval, requires an understanding of subjective and context-dependent fashion compatibility. However traditional methods, inspired by learning-to-rank approaches, have shown limited performance. To address these challenges, we propose the Compatibility-to-Retrieval Model (C2RM). First, we use Conditional Generative Adversarial Networks (cGANs) to generate target samples (templates) from seed items, providing conditioning signals for retrieval. Second, C2RM evaluates item-item and item-template compatibility, enhancing accuracy and reducing data requirements. Our extensive experiments on two datasets—FashionVC [6] and ExpFashion [7]—demonstrate C2RM's superior performance.

IIR2024: 14th Italian Information Retrieval Workshop, September 5-6, 2024, Udine, Italy

*An extended version of this work is currently under review.

†Corresponding author.

matteo.attimonelli@poliba.it (M. Attimonelli); claudio.pomo@poliba.it (C. Pomo); dietmar.jannach@aau.at (D. Jannach); tommaso.dinoia@poliba.it (T. Di Noia)

2. Related Work

Recent advances in fashion recommendation systems have significantly impacted e-commerce by enhancing user experience and pushing sales. This growth is partially driven by enhanced fashion item retrieval and recommendation research and advancing generative models.

Fashion Item Retrieval & Recommendation. Fashion recommendation systems, thanks to the economic benefits of e-commerce, focus on suggesting individual items or complete outfits. To offer personalized suggestions, fashion item recommendation [8, 9, 10] algorithms consider user preferences, behaviors, demographics, and context. Outfit recommendation extends this by creating cohesive ensembles, requiring a deep understanding of fashion aesthetics and trends. Complementary item retrieval [3, 4, 5], particularly top-bottom retrieval, is challenging due to the need for accurate compatibility modeling. Traditional approaches like BPR-DAE [6] and GP-BPR [11] often fail to generate novel yet compatible item combinations.

Generative Models for Top-Bottom Retrieval. Generative models have been integrated into retrieval systems to improve compatibility assessments. Notable examples include c+GAN [12] and CRAFT [13], focusing on generating compatible bottoms given a top. Advanced methods like DVBPR [14], FARM [15], and MGCM [16] leverage generative models to enhance retrieval performance, though they often overlook the quality of the generated items.

The C2RM model proposed in this work separates the training of generative and compatibility modules, optimizing each with distinct loss functions to achieve stable convergence and improved performance. This method advances the field by ensuring high-quality generation and accurate compatibility assessments.

3. Methodology

This section presents our proposed Compatibility Modeling strategy for fashion item retrieval, including two distinct phases. The first step is *Template Generation*: here we use an adapted Pix2Pix [17] architecture to generate templates of a compatible bottom item dress for a given input top dress, which is trained independently to perform an image-to-image translation task. The second step is *Compatibility Assessment*: using the generated template and the input top as a query, the C2RM model computes compatibility scores between the query and every other bottom clothes in the dataset. These scores facilitate the ranking and retrieval of potential matches.

Template Generation. This step aims to learn the hidden relations between the top and bottom distributions by adopting an image-to-image translation architecture to accomplish the conditioning generation task and learn a mapping between the source and target distributions. Specifically, the generative model is based on the Pix2Pix [17] architecture, which consists of two primary components: a U-Net generator and a PatchGAN discriminator. The U-Net generator acts as an autoencoder with skip-connections, which help to address the vanishing gradient problem [18] during training. On the other hand, the PatchGAN [17] discriminator produces a patch[1] of a pre-defined size rather than a single scalar output, assessing which patch of the image is realistic or fake.

[1] A patch is a specific portion of the image of a pre-defined size.

Compatibility-to-Retrieval Model. Our efforts focus on enhancing the quality of the generated samples and increasing the size of the generated template images to assess whether these improvements would also enhance compatibility modeling and the top-bottom retrieval tasks, as better and more realistic generated images are expected to increase retrieval performance. We designed the *C2RM* inspired by the work by Liu et al. [16]. C2RM leverages the bottoms generated by the cGAN, referred to as templates, for evaluating compatibility. Our model is built upon a pretrained ResNet [19] backbone to extract and project features from the top, bottom, and template images. To this end, we model compatibility as a combination of *item-item* and *item-template similarities*. The *item-item* score refers to the similarity between the latent representations of the top and candidate bottoms. Instead, the *item-template* compatibility focuses on the absolute distance between the latent representations of the template and candidate bottoms.

This approach ensures that item compatibility is evaluated accurately, enhancing retrieval performance. To maximize the compatibility between positive top-bottom pairs, we evaluate the compatibility scores as shown in Equation (1):

$$m_{ij} = \alpha \cdot \mathrm{sim}(t_i, b_j) + \beta \cdot \|b_j - s_i\|_1, \quad m_{ik} = \alpha \cdot \mathrm{sim}(t_i, b_k) + \beta \cdot \|b_k - s_i\|_1 \tag{1}$$

where t_i, b_j, and b_k belong to $\xi \coloneqq \{(i, j, k) \mid (t_i, b_j) \in \mathscr{P}, \quad b_k \in \mathscr{P} \setminus b_j\}$, and $\mathscr{P}$ represents the set of all compatible pairs in the catalog.

Thus, the loss function can be designed as Equation (2) where the strength of each component in the overall loss $\mathscr{L}$ is regulated by the hyperparameters γ and δ. Therefore, we define the BPR loss as $\mathscr{L}_{\text{bpr}}$ and the regularization component as $\mathscr{L}_{\text{reg}}$ with θ_c representing the parameters of the C2RM model.

$$\mathscr{L} = \gamma \cdot \mathscr{L}_{\text{bpr}} + \delta \cdot \mathscr{L}_{\text{reg}}, \quad \mathscr{L}_{\text{bpr}} = -\log \sigma(m_{ij} - m_{ik}), \quad \mathscr{L}_{\text{reg}} = \|\theta_c\|_2 \tag{2}$$

4. Experiments

We conducted a benchmarking study on the FashionVC [6] and ExpFashion [7] datasets[2] to evaluate our model against various baselines, namely BPR-DAE [6], MGCM [16], and Pix2PixCM [16]. Our main objective was to determine if the model meets the desired outcomes as per selected metrics.

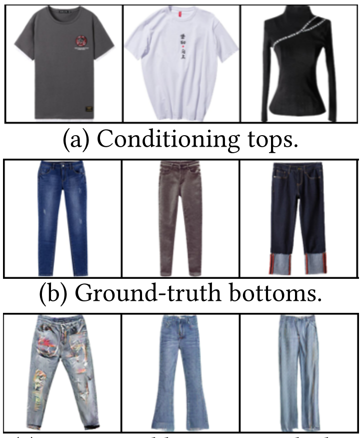
We employ standard metrics for this kind of purpose, namely Area Under the Curve (AUC) and Mean Reciprocal Rank (MRR). With the first one, we want to assess the performance of our compatibility model's step. We compute the AUC considering each top t_i with its positive bottom b_j and one negative bottom b_k. Regarding the MRR, we adopted it to showcase the ranking performance of our model compared to other baselines. We compute MRR

Table 1: Performance of the different models. **Boldface** and underlined indicate best and second-to-best values.

Model	FashionVC		ExpReduced	
	AUC	*MRR*	*AUC*	*MRR*
C2RM	**0.7209**	**0.0259**	**0.6952**	**0.0182**
MGCM	0.5939	0.0098	0.4674	0.0075
Pix2PixCM	0.5000	0.0080	0.5136	0.0078
BPR-DAE	<u>0.6590</u>	<u>0.0200</u>	<u>0.6457</u>	<u>0.0130</u>
RANDOM	0.4908	0.0081	0.4908	0.0063

[2]For dataset configuration and splitting we refer to Liu et al. [16].

by considering the position of the positive bottom b_j in the ranked list, where all other bottoms in the test set are treated as negatives. Table 1 displays the results from our benchmark setup, highlighting the performance of C2RM, which outperforms all other models across all datasets. Notably, BPR-DAE stands out as the second-best performer, showing a significant margin over other models on the FashionVC and ExpFashion datasets. This suggests simpler models such as BPR-DAE are highly competitive on smaller datasets like FashionVC and ExpFashion. This efficacy is further evidenced by the poor performance of MGCM and Pix2PixCM.



(a) Conditioning tops.

(b) Ground-truth bottoms.

(c) Generated bottoms with the proposed Generative Model.

Figure 1: Examples of the generation results using our cGAN.

Figure 1 illustrates the generation process of our model in detail. In this example, a t-shirt (top) is fed to the cGAN, generating a corresponding pant (bottom). The generated output demonstrates high quality and realism, significantly surpassing the results produced by models such as MGCM and Pix2PixCM[3]. This high-fidelity generation confirms that our approach effectively leverages the full information content and superior quality of the generated items, leading to notable improvements in compatibility, as measured by the AUC, and in retrieval performance, as indicated by the MRR. The enhanced generation quality underscores the robustness and effectiveness of our model in practical applications.

5. Conclusions

This study presents our innovative compatibility schema for fashion image retrieval, which utilizes paired image-to-image translation. Our model C2RM introduces two independently trained components designed for compatibility modeling. Despite being trained on a relatively small dataset, we demonstrate the reliability and stability of C2RM. Additionally, C2RM produces high-quality and realistic generations, outperforming baseline models in quality and realism.

For future work, we plan to introduce textual signals to enhance retrieval performance. Key areas of investigation will include determining the optimal stage for integrating these textual signals, whether at the generative model level or during the retrieval process. Additionally, we aim to incorporate personalization signals and explore the most effective methods for integrating these into our existing framework.

Acknowledgments

This work has been carried out while *Matteo Attimonelli* was enrolled in the Italian National Doctorate on Artificial Intelligence run by Sapienza University of Rome in collaboration with *Politecnico Di Bari*. This work was partially supported by the following projects: CT_FINCONS_III, OVS Fashion Retail Reloaded, LUTECH DIGITALE 4.0, VAI2C, IDENTITA, REACH-XY. We acknowledge the CINECA award under the ISCRA initiative, for the availability of high-performance computing resources and support.

[3]Due to space constraints, we show only our generated images. See [16] for a baseline comparison.

References

[1] Y. Deldjoo, F. Nazary, A. Ramisa, J. J. McAuley, G. Pellegrini, A. Bellogín, T. Di Noia, A review of modern fashion recommender systems, ACM Comput. Surv. 56 (2024) 87:1–87:37.

[2] M. Attimonelli, D. Danese, D. Malitesta, C. Pomo, G. Gassi, T. Di Noia, Ducho 2.0: Towards a more up-to-date unified framework for the extraction of multimodal features in recommendation, in: WWW (Companion Volume), ACM, 2024, pp. 1075–1078.

[3] K. Bibas, O. S. Shalom, D. Jannach, Semi-supervised adversarial learning for complementary item recommendation, in: WWW, ACM, 2023, pp. 1804–1812.

[4] J. Hao, T. Zhao, J. Li, X. L. Dong, C. Faloutsos, Y. Sun, W. Wang, P-companion: A principled framework for diversified complementary product recommendation, in: CIKM, ACM, 2020, pp. 2517–2524.

[5] J. Wu, J. A. S. Rodríguez, H. J. C. Pampín, Session-based complementary fashion recommendations, CoRR abs/1908.08327 (2019).

[6] X. Song, F. Feng, J. Liu, Z. Li, L. Nie, J. Ma, Neurostylist: Neural compatibility modeling for clothing matching, in: ACM Multimedia, ACM, 2017, pp. 753–761.

[7] Y. Lin, P. Ren, Z. Chen, Z. Ren, J. Ma, M. de Rijke, Explainable outfit recommendation with joint outfit matching and comment generation, IEEE TKDE 32 (2020) 1502–1516.

[8] R. He, J. J. McAuley, VBPR: visual bayesian personalized ranking from implicit feedback, in: AAAI, AAAI Press, 2016, pp. 144–150.

[9] Q. Liu, S. Wu, L. Wang, Deepstyle: Learning user preferences for visual recommendation, in: SIGIR, ACM, 2017, pp. 841–844.

[10] J. Chen, H. Zhang, X. He, L. Nie, W. Liu, T. Chua, Attentive collaborative filtering: Multimedia recommendation with item- and component-level attention, in: SIGIR, ACM, 2017, pp. 335–344.

[11] X. Song, X. Han, Y. Li, J. Chen, X. Xu, L. Nie, GP-BPR: personalized compatibility modeling for clothing matching, in: ACM Multimedia, ACM, 2019, pp. 320–328.

[12] S. Kumar, M. D. Gupta, c^{+}gan: Complementary fashion item recommendation, CoRR abs/1906.05596 (2019).

[13] C. P. Huynh, A. Ciptadi, A. Tyagi, A. Agrawal, CRAFT: complementary recommendations using adversarial feature transformer, CoRR abs/1804.10871 (2018).

[14] W. Kang, C. Fang, Z. Wang, J. J. McAuley, Visually-aware fashion recommendation and design with generative image models, in: ICDM, IEEE Computer Society, 2017, pp. 207–216.

[15] Y. Lin, P. Ren, Z. Chen, Z. Ren, J. Ma, M. de Rijke, Improving outfit recommendation with co-supervision of fashion generation, in: WWW, ACM, 2019, pp. 1095–1105.

[16] J. Liu, X. Song, Z. Chen, J. Ma, MGCM: multi-modal generative compatibility modeling for clothing matching, Neurocomputing 414 (2020) 215–224.

[17] P. Isola, J. Zhu, T. Zhou, A. A. Efros, Image-to-image translation with conditional adversarial networks, in: CVPR, IEEE Computer Society, 2017, pp. 5967–5976.

[18] R. Pascanu, T. Mikolov, Y. Bengio, On the difficulty of training recurrent neural networks, volume 28 of *JMLR Workshop and Conference Proceedings*, 2013, pp. 1310–1318.

[19] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: CVPR, IEEE Computer Society, 2016, pp. 770–778.