Fashion Compatibility Recommendation via Unsupervised Metric Graph Learning

Jiali Duan [α] **Xiaoyuan Guo** [β] **Son Tran** [γ] **C.-C. Jay Kuo** [α]

Abstract

In the task of fashion compatibility prediction, the goal is to pick an item from a candidate list to complement a partial outfit in the most appealing manner. Existing fashion compatibility recommendation work comprehends clothing images in a single metric space and lacks detailed understanding of users' preferences in different contexts. To address this problem, we propose a novel Metric-Aware Explainable Graph Network (MAEG). In MAEG, we propose an unsupervised approach to obtain representation of items in a metric-aware latent semantic space. Then, we develop a graph filtering network and Pairwise Preference Attention module to model the interactions between users' preferences and contextual information. Experiments on real world dataset reveals that MAEG not only outperforms the state-of-the-art methods, but also provides interpretable insights by highlighting the role of semantic attributes and contextual relationships among items.

1 Overview of our pipeline

An overview of our framework is shown in Figure 1, in which LSEN is trained unsupervised to extract metric-aware latent representations $x_i = [x_i^{p_1}; ...; x_i^{p_K}; x_i^g], i = 1, ..., N$, with K latent space and N items in this outfit. PPA models the user's preference by computing pair-wise attention over latent attributes, resulting in a weighted feature from K metric space. MAEG accounts for interactions between user's preference and contexts using graph filtering, trained in a multi-task setting (link prediction and compatibility prediction).

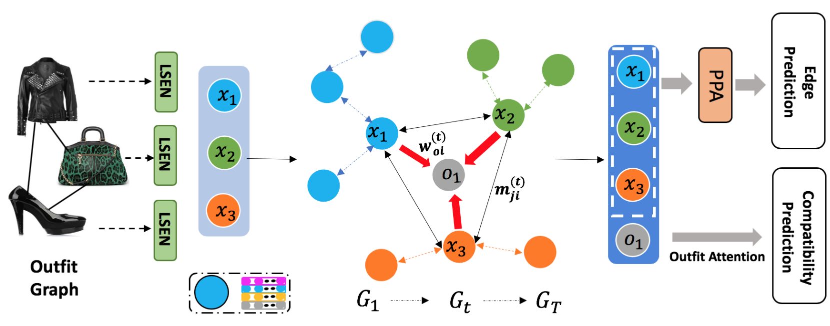

Figure 1: The architecture of Metric-Aware Explainable Graph Network (MAEG)

2 Experiments and Conclusion

Recommendation Performance: We evaluate our model on a large-scale real world dataset *Polyvore Outfits* [3] following the protocol in [3]. As shown in Table 1, our model obtains consistent improve-

[α]University of Southern California [β]Emory University [γ]Amazon

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

ment over state-of-the-art approaches [1, 2, 3]. To the best of our knowledge, we made the first attempt to demonstrate the effectiveness of leveraging metric-aware latent embedding in graph neural networks.

Method	FITB ACC	Compat. AUC
TA-CSN [3]	55.3%	0.86
SCE-Net [2]	61.6%	0.91
CA-GCN (wo/ ctx) [1]	43.3%	0.75
CA-GCN (w/ ctx) [1]	82.4%	0.99
Ours (wo/ ctx)	63.1%	0.93
Ours (w/ ctx)	86.7%	0.99
Ours+Outfit (w/ ctx)	**88.0%**	**0.99**

Table 1: Comparisons on FITB/Compatibility task over *Polyvore Outfits*.

Interpretability of Our Model: As shown in Figure 2, clothing items that are compatible in color, style, shape, etc are embed close to each other in our learned latent space and item that are more conspicuous seem to get more credits for the final "contributions", which to some extent coincides with human intuition. Finally, a user's favor over items or attributes tend to change given different contexts, which is captured during graph filtering, shown in Figure 3.

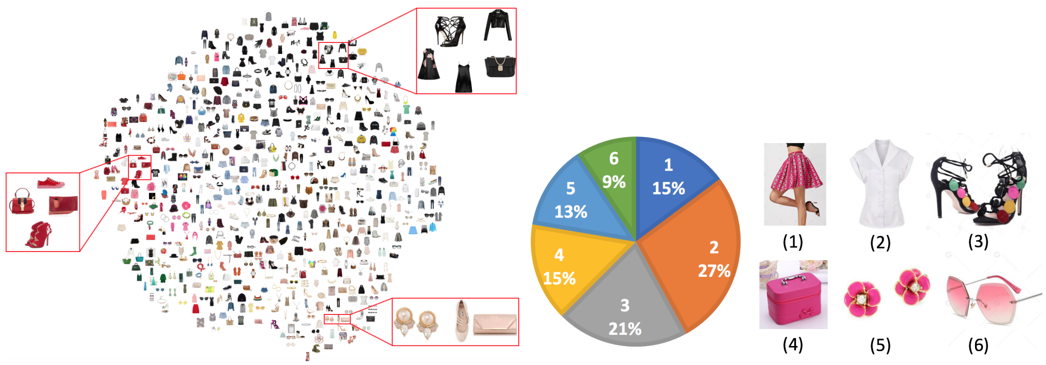

Figure 2: (left:) t-SNE plot of learned latent space; (right:) Attention distribution for each item

Figure 3: Qualitative results of our model for FITB prediction on *Maryland Polyvore (left)* and *Polyvore Outfits (right)*. Green box indicates the groundtruth and scores highlighted with red color are our predictions.

References

[1] Guillem Cucurull, Perouz Taslakian, and David Vazquez. Context-aware visual compatibility prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12617–12626, 2019.

[2] Reuben Tan, Mariya I Vasileva, Kate Saenko, and Bryan A Plummer. Learning similarity conditions without explicit supervision. *arXiv preprint arXiv:1908.08589*, 2019.

[3] Mariya I Vasileva, Bryan A Plummer, Krishna Dusad, Shreya Rajpal, Ranjitha Kumar, and David Forsyth. Learning type-aware embeddings for fashion compatibility. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 390–405, 2018.