Learning Personal Tastes in Choosing Fashion Outfits

Yusan Lin, Maryam Moosaei, Hao Yang
{yusalin,mmoosaei,haoyang}@visa.com
Visa Research
Palo Alto, CA, USA 94306

Abstract

With the emergence of fashion recommendation, many researchers have attempted to recommend fashion items that fit consumers' tastes. However, few have looked into fashion outfits as a whole when making recommendations. In this paper, we propose a neural network that learns one's fashion taste and predicts whether an individual likes a fashion outfit. To improve learning, we also develop a fashion outfit negative sampling scheme to sample fashion outfits that are different enough. With experiments on the collected Polyvore dataset, we find that using complete images of fashion outfits performs well when learning individuals' tastes toward fashion outfits. Our proposed negative sampling scheme also improves the model's performance significantly, compared to random negative sampling.

1. Introduction

Understanding individuals' fashion tastes can lead to significant benefits in multiple aspects. For the consumers, it helps pick their everyday outfits, and decide what apparel items to purchase based on their fashion tastes. For the merchants, it boosts their revenues by adding recommendation modules to their online or physical stores. However, fashion taste is a complicated concept that is hard to grasp. Recently, a new service has been surfacing in the fashion industry: the recommendation system. Personal shopping services such as Affinity and StitchFix recommend consumers what fashion items to purchase based on their own tastes. Both have shown great popularity and success.

Making high-quality fashion recommendation based on individuals' fashion tastes involve two major challenges. First of all, it is more meaningful to recommend individuals with a set of items that make a fashion outfit, rather than individual items. Secondly, many data sources for learning fashion tastes from consist of extremely sparse connections between individuals and the fashion outfits they liked/created/purchased, which lead to data sparsity problem. In this paper, we focus on these two challenges and propose a neural network model to handle them. To encounter the above challenges, we propose to directly use outfit images that combine all the fashion items for our neural network to learn, and design a new negative sampling scheme to select outfits that individuals are unlikely to like based on what they liked before. Contributions of this paper are:

1. We leverage fashion data from Polyvore and propose a neural network model that, given a fashion outfit and a user, predicts whether the user likes the fashion outfit. The learned model can be used to further recommend fashion outfits to individuals based on the learned personalized outfit scores.

2. We propose a new way of negative sampling that reflects well the fashion outfits unlikely to be liked by an individual. Our proposed scheme encounters the data sparsity issue and is shown to significantly improve the model's performance compared to randomly selecting negative samples.

2. Related Work

Most of the available fashion recommendation systems use keyword search [5], purchased histories [6], and user ratings [3, 4] to recommend items to individuals. These methods do not consider the visual appearance of items, which is a key feature in a visual domain like fashion. To address this limitation, many have focused on developing content-based methods. For example, [3] developed a deep neural network that combines implicit feedback from users with visual content to model individuals' fashion tastes and recommend items that maximize a user's preference score. However, many times users are interested in an entire outfit with multiple items that meet their personal tastes. Recommending personalized fashion outfits is more complex than recommending single items. In this paper, we present our ongoing research in developing a personalized outfit recommendation system. Our system learns user preferences in outfit selection from their online activities on Polyvore.

3. Proposed Methodology

3.1. Fashion Outfit Preference Prediction Model

To learn individuals' tastes toward fashion outfits, we formulate the research problem as follows. Let O be all the fashion outfits, and U be all the users. Given a fashion outfit o and a user u, our goal is to predict whether user u likes fashion outfit o.

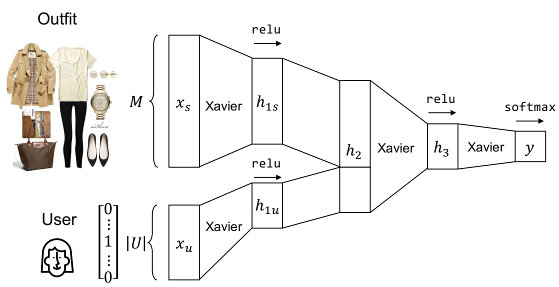

Figure 1. Fashion outfit recommendation neural network model.

To approach the problem, we propose a neural network model with structure design shown in Figure 1. The network takes two inputs: an outfit o and a user u. We encode u into an one-hot encoding vector with dimension $|U|$ as x_u. We then pass o to a CNN network to generate a representation of dimension M, denoted as x_s. x_s and x_u are then passed through fully-connected layers (FC layers) with weights W_{1s} and W_{1u}, and become lower-dimensional representations h_{1s} and h_{1u}, respectively. Next, we concatenate h_{1s} and h_{1u} to get h_2. h_2 is then passed through another FC layer with weights W_{23} to generate h_3, which represents a joint embedding. Afterwards, h_3 is passed through another FC layer with weights W_{3y} and the output of this FC layer is input to a softmax layer which generates a score $\hat{y}$ between 0 and 1. A score of 0 means that user u does not like outfit o, and a score of 1 means that user u likes outfit o. Note that this model is a binary classifier, and all of the activation functions are relu. $\hat{y}$ can then be expressed as follows.

$$\hat{y} = \texttt{softmax}\big(h_3 \cdot W_{3y}\big) \tag{1}$$

The objective of the model is to minimize the loss:

$$\mathcal{L}(\Theta) = \sum_{i}^{N} \left(\hat{y}_i - y_i\right)^2 \tag{2}$$

where $\Theta = \{W_{s1}, W_{u1}, W_{23}, W_{3y}\}$ is the set of model parameters, and N is the number of instances to learn.

3.2. Fashion Outfit Negative Sampling

Let the dataset be $D_{pos} = \langle u, o_i \rangle$, where each $\langle u, o_i \rangle$ is a positive sample of user-liking-outfit. One naive way to select negative samples is, for each $\langle u, o_i \rangle \in D_{pos}$, randomly select a o_j, where $\langle u, o_j \rangle \notin D_{pos}$. However, just because u has never liked o_j before does not mean u will not like it in the future. Therefore, we propose another scheme for selecting negative fashion outfits that are at least λ-*different*. Given $\langle u, o_i \rangle \in D_{pos}$, in the space of fashion outfit features, we randomly pick o_j and use a distance measure of choice (e.g., cosine distance) to measure its distance to o_i, $d(o_i, o_j)$. If $d(o_i, o_j) < \lambda$, we re-select a negative sample and re-check its distance to o_i. Otherwise, we add $\langle u, o_j \rangle$ to D_{neg}. We repeat the steps above for all $\langle u, o \rangle \in D_{pos}$ until $|D_{pos}| = |D_{neg}|$.

Metric	Random	λ- differen
loss	2.025	0.580
accuracy	0.523	0.718
precision	0.376	0.758
recall	0.383	0.642
f1	0.352	0.695

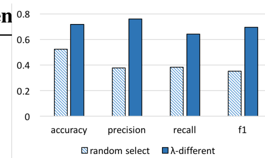

Table 1. Model performance.

4. Evaluation

We created a dataset by crawling Polyvore, a fashion social network allowing users to create fashion outfits from different items, or like and save outfits created by others.We crawled profiles of 150 randomly selected users to create a dataset of 66K fashion outfits and 158,503 fashion items.

We converted each outfit to a feature vector using a DenseNet, pre-trained on ImageNet [2, 1]. We also added batch normalization and drop-out (rate = 50%) to each layer of our network and used Xavier initialization for all the FC layers. Through empirical testing, we found the best configuration is $|h_{1s}| = 256$, $|h_{1u}| = 16$, and $|h_3| = 16$. To select negative samples for each user, based on our proposed negative sampling scheme in Section 3, we picked the 75$^{\text{th}}$ quantile as λ, which is 441.4.

We compare the performance of the model using two negative sampling approaches: random sampling and λ-different sampling. The results are shown in Table 1. As shown, the λ-different negative sampling leads the model to better performance compared to when sampling random negative samples.

5. Conclusion

In this paper, we presented our ongoing research on developing a personalized fashion outfit recommendation system. We developed a neural network model that takes a fashion outfit and an individual as inputs and predicts whether the individual likes the given fashion outfit. We found that using fashion outfit images as a whole is very informative for the network which eliminates the need to carefully detecting and segmenting every single item in the outfit. Our preliminary results show that our proposed approach can accurately predict whether an individual likes a fashion outfit.

References

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009. 2

[2] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 2261–2269, 2017. 2

[3] W.-C. Kang, C. Fang, Z. Wang, and J. McAuley. Visually-aware fashion recommendation and design with generative image models. *arXiv preprint arXiv:1711.02231*, 2017. 1

[4] X. Qian, H. Feng, G. Zhao, and T. Mei. Personalized recommendation combining user interest and social circle. *IEEE Transactions on Knowledge and Data Engineering*, 26(7):1763–1777, 2014. 1

[5] K. Vaccaro, S. Shivakumar, Z. Ding, K. Karahalios, and R. Kumar. The elements of fashion style. In *29th Annual Symposium on User Interface Software and Technology*, 2016. 1

[6] J. Wang, B. Sarwar, and N. Sundaresan. Utilizing related products for post-purchase recommendation in e-commerce. In *ACM Conference on Recommender Systems*, 2011. 1