SMART FASHION RECOMMENDER APPLICATION

Team ID: PNT2022TMID34679

Team Size: 4

Team Leader: SHIJU JERCY J (962219104103)

Team member: LIFERIN JENI B (962219104701)

Team member: RAJA GOWSHIKA V (962219104083)

Team member: NIDHIN R S (962219104079)

Smart Fashion Recommender Application

With an increase in the standard of living, peoples' attention gradually moved towards fashion that is concerned to be a popular aesthetic expression. Humans are inevitably drawn towards something that is visually more attractive. This tendency of humans has led to the development of the fashion industry over the course of time. However, given too many options of garments on the e-commerce websites, has presented new challenges to the customers in identifying their correct outfit. Thus, in this project, we proposed a personalized Fashion Recommender system that generates recommendations for the user based on an input given. Unlike the conventional systems that rely on the user's previous purchases and history, this project aims at using an image of a product given as input by the user to generate recommendations since many-a-time people see something that they are interested in and tend to look for products that are similar to that. We use neural networks to process the images from Fashion Product Images Dataset and the Nearest neighbour backed recommender to generate the final recommendations.

Introduction

Humans are inevitably drawn towards something that is visually more attractive. This tendency of humans has led to development of fashion industry over the course of time. With introduction of recommender systems in multiple domains, retail industries are coming forward with investments in latest technology to improve their business. Fashion has been in existence since centuries and will be prevalent in the coming days as well. Women are more correlated with fashion and style, and they have a larger product base to deal with making it difficult to take decisions. It has become an important aspect of life for modern families since a person is more often than not judged based on his attire. Moreover, apparel providers need their customers to explore their entire product line so they can choose what they like the most which is not possible by simply going into a cloth store.

Related work

In the online internet era, the idea of Recommendation technology was initially introduced in the mid-90s. Proposed CRESA that combined visual features, textual attributes and visual attention of the user to build the clothes profile and generate recommendations. Utilized fashion magazines photographs to generate recommendations. Multiple features from the images were extracted to learn the contents like fabric, collar, sleeves, etc., to produce recommendations. In order to meet the diverse needs of different users, an intelligent Fashion recommender system is studied based on the principles of fashion and aesthetics. To generate garment recommendations, customer ratings and clothing were utilized in the history of clothes and accessories, weather conditions were considered in to generate recommendations.

Proposed methodology

In this project, we propose a model that uses Convolutional Neural Network and the Nearest neighbour backed recommender. Initially, the neural networks are trained and then an inventory is selected for generating recommendations and a database is created for the items in inventory. The nearest neighbour’s algorithm is used to find the most relevant products based on the input image and recommendations are generated.

Training the neural networks

Once the data is pre-processed, the neural networks are trained, utilizing transfer learning from ResNet50. More additional layers are added in the last layers that replace the architecture and weights from ResNet50 in order to fine-tune the network model to serve the current issue.

Getting the inventory

The images from Kaggle Fashion Product Images Dataset. The inventory is then run through the neural networks to classify and generate embeddings and the output is then used to generate recommendations.

Recommendation generation

To generate recommendations, our proposed approach uses Sklearn Nearest neighbours Oh Yeah. This allows us to find the nearest neighbours for the given input image. The similarity measure used in this Project is the Cosine Similarity measure. The top 5 recommendations are extracted from the database and their images are displayed.

Experiment and results

The concept of Transfer learning is used to overcome the issues of the small size Fashion dataset. Therefore, we pre-train the classification models on the DeepFashion dataset that consists of 44,441 garment images. The networks are trained and validated on the dataset taken. The training results show a great accuracy of the model with low error, loss and good f-score.

Dataset Link

Kaggle Dataset Big size 15 GB

Kaggle Dataset Small size 572 MB

Installation

Use pip to install the requirements.

```
pip install -r requirements.txt
```

Usage

To run the web server, simply execute streamlit with the main recommender app:
streamlit run main.py

Built With

OpenCV - Open-Source Computer Vision and Machine Learning software library
Tensorflow - TensorFlow is an end-to-end open-source platform for machine learning.

Tqdm - tqdm is a Python library that allows you to output a smart progress bar by wrapping around any iterable.

streamlit - Streamlit is an open-source app framework for Machine Learning and Data Science teams. Create beautiful data apps in hours, not weeks.

pandas - pandas is a fast, powerful, flexible and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language.

Pillow - PIL is the Python Imaging Library by Fredrik Lundh and Contributors.

scikit-learn - Scikit-learn is a free software machine learning library for the Python programming language.

opencv-python - OpenCV is a huge open-source library for computer vision, machine learning, and image processing.

Major Challenges

In this section we will describe the major challenges faced by recommender systems in the fashion domain.

1.Fashion item representation:

Traditional recommender systems such as Collaborative Filtering or Content-Based Filtering have difficulties in the fashion domain due to the sparsity of purchase data, or the insufficient detail about the visual appearance of the product in category names [123]. Instead, more recent literature has leveraged models that capture a rich representation of fashion items through product images [51, 71], text descriptions or customer reviews [28, 164], or videos [153] which are often learned through surrogate tasks like classification or product retrieval. However, learning product representations from such input data requires large datasets to generalize well across different image (or text) styles, attribute variations, etc. Furthermore, constructing a representation that learns which product features customers take most into account when evaluating fashion products is still an open research problem.

2.Fashion item compatibility:

Training a model that is able to predict if two fashion items 'go together,' or directly combine several products into an outfit, is a challenging task. Different item compatibility signals studied in recent literature include co-purchase data [107, 139], outfits composed by professional fashion designers [48], or combinations found by analysing what people wear in social media pictures [72, 131]. From this compatibility information, associated image and text data is then used to learn to generalize to stylistically similar products. Some works explicitly model the latent style types [19]. An additional under explored difficulty for compatibility prediction is the dependency on trends, seasonality, location or social group. Current approaches usually leverage image and text information.

3.Personalization and fit:

The best fashion product to recommend depends on factors such as the location where the outfit will be used [8, 23, 80, 143], the season or occasion [98, 105, 162], or the cultural and social background of the customer [81, 131, 165]. A challenging task in fashion recommendation systems is how to discover and integrate these disparate factors [127, 147]. Current research often tackles these tasks by utilizing large-scale social media data. As discussed earlier, a personalization dimension very particular to the fashion domain is that of fit. In addition to predicting what size of a product will be more comfortable to wear, body shape can influence stylistic choices [58, 60, 122].

4.Interpretability and Explanation:

Most of the existing fashion recommender systems in the literature focus on improving predictive performance, treating the model as a black box. However, deploying accountable and interpretable systems able to explain their recommendations can foster user loyalty in the long term and improve the shopping experience. Current models generally offer explanations through highlighted image regions and attributes or keywords [47, 59, 89, 110, 149, 155].

5.Discovering Trends:

Being able to forecast consumer preferences is valuable for fashion designers and retailers in order to optimize product-to-market fit, logistics and advertising. Many factors are confounded in what features are considered ‘fashionable’ or ‘trendy’, like seasonality [105], geographical influence [8], historical events [63] or style dynamics [9, 44, 100]. Again, social media is a useful resource leveraged by researchers [41, 67].

The challenges above, as well as many other issues have been discussed in the studied research works, and they are reviewed in the following sections. For instance, visual modelling of fashion items (cf. Section 2.2.2), modelling of fashion outfits (cf. Section 2.1.2), geo-temporal localization (cf. Section 2.2.3), attribute prediction for fashion and leveraging multi-modal data (cf. Section 2.2.2), explain ability in fashion recommendation (cf. Section 2.1.4) and learning style (cf. Section 2.2.2).

Search strategy for relevant papers

We primarily relied on papers indexed in DBLP,1 a prominent computer science bibliographic database, to identify publications that comprise the state of the art in fashion item and outfit recommendation. Our search approach was divided into two stages: finding relevant publication collections; and (ii) post-processing and filtering the final list. The total number of articles analysed exceeds 50, with the majority of recognized research papers published between 2014 and 2021, an indication of the topic’s originality and freshness. While we do not claim that this study is exhaustive in terms of the collected papers, we believe it gives a comprehensive overview of current achievements, trends, and what we deem to be the most significant challenges and tasks in fashion recommender systems. Additionally, it gives valuable information to both industry practitioners and academics.

Related Surveys

To put this survey in context, we identified and present related review and survey articles to explain in which ways our article differs from and extends earlier work. In a recent work, Deldjoo et al. [39] presented a survey of RS leveraging multimedia content, i.e., visual, audio, and/or textual features. The domains studied in this survey include various ones such as media streaming for audio and video recommendation, e-commerce for recommending different products including fashion items, news, and information recommendation, social media, and so forth. While fashion RS were also discussed, the authors only included a small portion of the topics and papers in this domain. Here, we discuss and present a comprehensive survey of significant tasks, challenges, and types of content used in the fashion RS field.

We have also identified surveys [29, 170] where the authors present a literature review of techniques at the intersection of fashion and computer vision (CV) and/or natural language processing (NLP). While we find these works relevant to this article, they remain largely different from the review presented here as those systems are not focused on RS but on other aspects of the fashion domain, such as text generation from images or pose estimation. Moreover, as another point of difference, we also provide recent techniques dealing with item visual and textual content representation exploited by RS approaches.

Perhaps the most relevant work to our current survey is a recent book chapter by Jaradat et al. [75] on fashion RS. This chapter focuses on discussing the state of the art of fashion recommendation systems; in particular, the authors affirm that deep learning represented a turning point with respect to the canonical approaches and therefore the authors examined four different tasks that use this new approach. Additionally, they provided examples and possible problems and their evaluation. In particular, the authors focused their review on tasks related to social media and the size recommendation problem (see Section 2.1.3, where we introduce this task in detail). In our survey, in addition to analysing the state of the art of the most commonly used algorithms in a wide range of tasks, we went in depth to understand which are the main features used by the more modern fashion recommender systems. In fact, an extensive discussion is held on how both the user and the items, with their characteristics, can be a source for the definition of models with accurate.

Literature

McAuley et al. [1] devised a parametric distance transformation that assigns a lower distance to garment pairings that fit well than to those that do not. And provided Image-based recommendations on styles and substitutes.

Hu et al. [2] conducted a preliminary investigation into personalised outfit recommendation. To describe the user-item and item-item interactions, a functional tensor factorization method was presented. They proposed A functional tensor factorization approach.

Veit et al. [4] learned feature transformation for a compatibility measure between pairs of objects using a Siamese CNN architecture. All of these works focused solely on the compatibility of two things. Furthermore, they simply modelled broad matching criteria and ignored the issue of personalisation.

Thombre in [3] used image segmentation and Kalman filter to realize Human detection and tracking. Orrite-Urunuela proposed a statistical model for detection and tracking of human silhouette and the corresponding 3D skeletal structure in gait sequences [5].

How-Lung [6] provided an outdoor aquatic surveillance system for human motion tracking and detection.

Ajmani et al. [7] present a novel method for content-based recommendation of media-rich commodities with the use of probabilistic multimedia ontology. Proposed an ontology based personalized garment recommendation system.

Li et al. [8] utilized the HMM of recommended items to match customers' model according to customer data. The second method is the collaborative filtering-based recommendations algorithm. Proposed Content-Based Filtering Recommendation Algorithm.

For instance, Nogueira et al. [9] presented a new collaborative filtering strategy that utilizes the visual attention to characterize images and alleviate the new item cold-start problem. The rule-based recommendation algorithm is the third method.

Hwang et al. [10] put forward a method to generate the automatic rules with the user's items and made a suggestion on the best rule. The fourth method is the utility-based recommendation.

For instance, Scholz et al. [11] found that exponential utility functions are better geared to predicting optimal recommendation ranks for products, and linear utility functions perform much better in estimating customers' willingness.

Koenig in [12] developed a system toward real-time human detection and tracking in diverse environments. However, mostly the researchers focus on the point of human detection and tracking in complex scene, while refined contour extraction of human in dynamic scene is still an open question.

Categorization of Fashion Recommender System

We categorize fashion recommender systems according to their task (cf. Section 2.1), and the input data they use to perform that task (cf. Section 2.2). We will discuss these two categorizations in more detail in the next sections.

1.Categorization based on task:

By task we refer to the internal goal the fashion RS aims to achieve. This affects, in particular, the expected output of the algorithms. We identified five main tasks in the academic literature that fashion RS aim to achieve: (i) Fashion item recommendation, (ii) Fashion pair and outfit recommendation, (iii) Size recommendation, (iv) Explanation for Fashion Recommendation and (v) Other fashion prediction tasks. There are a few other tasks that so far have not gathered much attention, but are growing in popularity; they will be summarized later under the general subsection (v).

We discuss these three categories by presenting a broad definition of each task formally, while leaving a more detailed discussion of several prominent research works for Section 3

Fashion item recommendation: The fashion item recommendation task, similar to the classical recommendation problem, focuses on suggesting individual fashion items (clothing), that match users' preferences. Definition 2.1 (Fashion item recommendation). Let U and I denote a set of users and fashion items in a system, respectively. Each user $u \in$ U is related to I + u, the set of items she has consumed.

Given a utility function g: U × I → R the Item Recommendation Task is defined as

$$\forall u \in U,\ i* u = \text{argmax}\ i \in I \backslash I+ u\ g\ (u, i)$$

where $i * u$ denotes the best matching item not consumed by the user u before. The preference of user u on item i could be encoded as $sui \in$ S, a continuous-value score (e.g., 1-5 Likert scale), or implicit feedback in which we assume the user likes the item if she has interacted with (i.e., reviewed, purchased, clicked) the item (i.e., sui = 1). I + u represents the set of (u, i) pairs for which sui is known. The task of personalized Top-K fashion item recommendation problem is formally defined as identifying, for user u, a set of ranked lists of items Xu = {$i1$, $i2$, ..., iK} that match user preference.
A simple yet effective pure CF model that serves as a foundation for many model-based VRS is BPR-MF [119]. Given a user u, and an item i, the core predictor in this model is given according to:

$$s\hat{}u, i = p\ T\ u\ q_i$$

where pu, $qi \in$ R F are the embedding vectors for user u and item i, respectively, and F is the size of the embedding vector. This predictor is known as Matrix Factorization (MF), the parameters of the model can be learnt using Bayesian Personalized Ranking (BPR), a pairwise ranking optimization framework. We will show later in Section 3.1 how different authors have used this model as a starting point in the fashion domain, so that it is extended to consider other types of inputs and learning scenarios.

Fashion pair and outfit recommendation: Fashion outfits are sets of N items that are worn together, e.g., for an outdoor wedding, graduation party, baby shower, and so forth. The simplest form of a fashion outfit is when N = 2, i.e., two different items that look good when paired together, such as an orange shirt and a blue pair of jeans. However, in general outfits in online fashion store can be composed of items in different categories (e.g., show, bottom, top, hat, bag) that share some stylistic relationship.

Definition 2.2 (Fashion outfit composition). Let O = {$i1$, $i2$, ..., iN} ∈ O denote a fashion outfit composed of a set of compatible fashion items, in which $in \in$ I am the n-the fashion item in the outfit O, and O is the set of all possible outfits, N is the length of the outfit, whose value is not fixed and can change with each outfit. Fashion Outfit Composition is formulated as follows: "given a scoring function $s(O)$ that indicates how well the outfit $O \in$ O is composed, find an outfit O that maximizes this utility":

$$O * = \text{argmax}\ s(Oj)$$
$$Oj \in O$$

where s is the outfit utility function, a scoring function that takes into account different types of relationships between fashion items to generate an outfit compatibility score. It is worth noting that, given this general task definition, the evaluation of fashion outfit composition is typically performed via fill-in-the-blank (FITB) or outfit compatibility score prediction [113] described in Section 4.

Moreover, it is possible to encode several objectives relevant to the fashion domain in the definition of the outfit composition scoring function by incorporating domain knowledge. For instance, McAuley et al. [107] defined compatibility according to complementary relationships (e.g., how much a white shirt complements blue pants), and similarity (how much one item in the outfit is visually similar to another item). Hsiao and Grauman [62] model the outfit scoring function $s(Oj)$ as the superposition of two objectives:

$$O * j = \underset{Oj \in \mathrm{O}}{\operatorname{argmax}}\ c(Oj) + v(Oj)$$

where $c(.)$ and $v(.)$ denote the compatibility score (how much pairs complement each other) and versatility score (defined as coverage of all styles), respectively. Parameters of the outfit scoring functions may be personalized to each style/user, and learned from user interaction data. As an example, in [93] the authors propose a system where compatibility is computed according to the image and category of every item in the outfit.

2.Categorization based on input:

By input, we refer to the data fashion RS use to train their models. They could be based on a U-I matrix together with side information beyond the U-I matrix (related to users and item), and contextual information. This information determines the type of RS, according to CF, hybrid (CF+CBF) systems, and context-aware (CA) systems. In Table 3, we provide a classification of data/computational features used for the design of fashion RS by authors over the years.

Generative Fashion Recommendation Models

Most conventional RS are not suitable for application in the fashion domain due to unique characteristics hidden in this domain. Recently, GAN-based models have been used for fashion generation and fashion recommendation with promising performance. GANs gain their power exploiting their generative power allowing them to synthesize realistic-looking fashion items Deldjoo et al. [38]. This aspect can inspire the aesthetic appeal/curiosity of customers and designers and motivate them to explore the space of potential fashion styles.

[CRAFT] Huynh et al. [68] address the problem of recommending complementary fashion items based on visual features by using an adversarial process that resembles a GAN and uses a conditional feature transformer as G and a discriminator D. One main distinction between this work and the prior literature is that the ⟨input, output⟩ pair for G are both features (here features are extracted using pre-trained CNNs [132]), instead of ⟨image, image⟩ or hybrid types such as ⟨image, features⟩ explored in numerous previous works [141, 168]. This would allow the network to learn the relationship between items directly on the feature space, spanned by the features extracted. The proposed system is named complementary recommendation using adversarial feature transform (CRAFT) since in the model, G acts like a feature transformer that—for a given query product image q—maps the source feature sq into a complementary target feature $t\hat{}q$ by playing a min-max game with D with the aim to classify fake/real features. For training, the system relies on learning the co-occurrence of item pairs in real images. In summary, the proposed method does not generate new images; instead, it learns how to generate features of the complementary items conditioned on the query item.

[DVBPR] Deep visual Bayesian personalized ranking (DVBPR) [79] is one of the first works to exploit the visual generative power of GANs in the fashion recommendation domain. It aims to generate clothing images based on user preferences. Given a user and a fashion item category (e.g., tops, t-shirts, and shoes), the proposed system generates new images—i.e., clothing items—that are consistent with the user's preferences. The contributions of this work are two-fold: first, it builds an end-to-end learning framework based on the Siamese-CNN framework. Instead of using the features extracted in advance, it constructs an end-to-end system that turns out to improve the visual representation of images. Second, it uses a GAN-based framework to generate images that are consistent with the user's taste. Iteratively, G learns to generate a product image integrating a user preference maximization objective, while D tries to distinguish generated images from real ones. Generated images are quantitatively compared with real images using the preference score (mean objective value), inception score [121], and opposite SSIM [108]. This comparison shows an improvement in preference prediction in comparison with non-GAN based images. At the same time, the qualitative comparison demonstrates that the generated images are realistic and plausible, yet they are quite different from any images in the original dataset—they have standard shape and colour profiles, but quite different styles.

[MrCGAN] Shih et al. [126] propose a compatibility learning framework that allows the user to visually explore candidate compatible prototypes (e.g., a white t-shirt and a pair of blue-jeans). The system uses metric-regularized conditional GAN (MrCGAN) to pursue the item generation task. It takes as input a projected prototype (i.e., the transformation of a query image in the latent "compatibility space"). It produces as the output a synthesized image of a compatible item (the authors consider a compatibility notion based on the complementary of the query item across different catalogue categories). Similar to the evaluation protocol in [68], the authors conduct online user surveys to evaluate whether their model could produce images that are perceived as compatible. The results show that MrCGAN can generate compatible and realistic images under compatibility learning settings compared to baselines.

[Yang et al. & c +GAN] Yang et al. [157] address the same problem settings of MrCGAN [126] by proposing a fashion clothing framework composed of two parts: a clothing recommendation model based on BPR combined with visual features and a clothing complementary item generation-based GAN. Notably, the generation component takes as input a piece of clothing recommended in the recommendation model and generates clothing images of other categories (i.e., tops, bottom, or shoes) to build a set of complementary items. The authors follow a similar qualitative and quantitative evaluation procedure as DVBPR [79] and further propose a compatibility index to measure the compatibility of the generated set of complementary items. A similar approach has also been proposed in c +GAN [84], to generate a bottom fashion item paired with a given top.

Other Fashion Recommender System Algorithms

Depending on the task, fashion item and outfit recommendation, and size recommendation as we identified in Section 2.1, a variety of other models have been used in the studied research work. For example, RNNs [21, 92], graph-modelling [27, 87, 107], two-input (Siamese) CNNs [79, 113, 158, 164, 167], using attention mechanisms [92] and so forth. Detailed discussion on these approaches is left as a future direction.

References

[1]. J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, "Image-based recommendations on styles and substitutes," in Proceedings of the International ACM Conference on Research and Development in Information Retrieval (SIGIR), 2015.

[2]. Y. Hu, X. Yi, and L. S. Davis, "Collaborative fashion recommendation: A functional tensor factorization approach," in Proceedings of the 23rd Annual ACM Conference on Multimedia Conference, 2015.

[3]. J. L. D. Thombre, D.V.; Nirmal, "Human detection and tracking using image segmentation and kalman filter," Intelligent Agent Multi-Agent Systems, pp. 1 - 5, 2009.

[4]. A. Veit, B. Kovacs, S. Bell, J. McAuley, K. Bala, and S. Belongie, "Learning visual clothing style with heterogeneous dyadic co-occurrences," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015.

[5]. J. H.-J. J. R. G. Orrite-Urunuela, C.; del Rincon, "2d silhouette and 3d skeletal models for human detection and tracking," 17th International Conference on Pattern Recognition, vol. 4, pp. 244 - 247, 2004.

[6]. A. J. W. W.-Y. Y. L. J. How-Lung Eng; Kam, "Human detection and tracking within hostile aquatic environments," 12th International Conference on Image Analysis and Processing, pp. 133 - 138, 2003.

[7]. S. Ajmani, H. Ghosh, A. Mallik, and S. Chaudhury, "An ontology based personalized garment recommendation system," in Proceedings of the IEEE/WIC/ACM International Joint Conference onWeb Intelligence and Intelligent Agent TechnologyWorkshops, WI-IATW2013, pp. 17-20, November 2013.

[8]. H. Li, F. Cai, and Z. Liao, "Content-Based Filtering Recommendation Algorithm Using," in Proceedings of the HMM, pp. 275- 277, 2012.

[9]. E. A. Nogueira, E. V. De Melo, E. R. De Faria, and D. Guliato, "IKB-MS: A collaborative filtering approach associated with human visual attention for clothing recommendation," in Proceedings of the 21st Brazilian Symposium on Multimedia and the Web, WebMedia 2015, pp. 149-156, October 2015.

[10]. I. Hwang, M. Kim, and H. J. Ahn, "Data pipeline for generation and recommendation of the IoT rules based on open text data," in Proceedings of the 30th IEEE International

Conference on Advanced Information Networking and ApplicationsWorkshops, WAINA 2016, pp. 238-242, March 2016.

[11]. M. Scholz, V. Dorner, M. Franz, and O. Hinz, “Measuring consumers’ willingness to pay with utility-based recommendation systems,” Decision Support Systems, vol. 72, pp. 60-71, 2015.

[12]. N. Koenig, “Toward real-time human detection and tracking in diverse environments,” IEEE 6th International Conference on Development and Learning, pp. 94 - 98, 2007.

Conclusion

In this project, we have presented a novel framework for fashion recommendation that is driven by data, visually related and simple effective recommendation systems for generating fashion product images. The proposed approach uses a two-stage phase. Initially, our proposed approach extracts the features of the image using CNN classifier i.e., for instance allowing the customers to upload any random fashion image from any E-commerce website and later generating similar images to the uploaded image based on the features and texture of the input image. It is imperative that such research goes forward to facilitate greater recommendation accuracy and improve the overall experience of fashion exploration for direct and indirect consumers alike.