DEEP LEARNING BASED CLOTHING MATCHING: A SURVEY

Cuijiao Zhang
Tiangong University

ABSTRACT

In modern society, clothing has gradually become an important beauty-enhancing product, playing an important role in human’s social life. Nevertheless, each outfit is composed of items of many different types (e.g. top, bottom, shoes) that share some complicated relationship with one another, how to make a suitable outfit has become a daily headache for many people, especially those who do not have much sense of aesthetics. In recent years, the booming fashion sector and its huge potential benefits have attracted tremendous attention from many research communities. Moreover, deep learning has garnered considerable recognition by researchers in many research fields such as computer vision and natural language processing, owing not only to stellar performance but also the attractive property of learning feature representations from scratch. The influence of deep learning is also pervasive, recently demonstrating its effectiveness when applied to information retrieval and outfit recommender systems research. In the light of this, many research efforts have been dedicated to the task of complementary clothing matching and have achieved great success relying on the advanced data-driven neural networks. However, finding suitable architectures, the selection and extraction of clothing features is a challenge for researchers. This article aims to provide a comprehensive review of recent research efforts on deep learning based clothing matching. More concretely, we first review the research on the clothing matching system and categorized deep learning-based clothing matching methods into two categories, namely collaborative method and content-based method. We then carry out a detailed introduction to this relevant research on deep learning based clothing matching.

KEYWORDS: Clothing Matching; Deep Learning; Survey.

1. INTRODUCTION:
With the influence of fashion magazines and fashion industries going online, clothing fashions are attracting more and more attention. According to the Fashion United, the global fashion and apparel industry is valued at three trillion dollars, making up two percent of the world's gross domestic product. The blossoming of the fashion market demonstrates people's great demand of clothing. Clothing matching refers to coordinating complementary fashion items such as the tops and bottoms in styles and colors to achieve a decent and generous effect on the whole. In fact, clothing matching has become an indispensable part of people's daily life, since a properly coordinated outfit can improve one's appearance greatly.

Clothing Matching, however, is a complex task that depends on subjective notions of style, context, and trend – all properties that may vary from one individual to another and evolve over time. Obviously, clothes matching is a subjective issue rather than objective. so, it is impossible to simply and directly calculate the degree of compatibility. In addition, there are various kinds of factors involving style, color, material and other factors, and the relationship among these factors is quite complicated. Therefore, the clothing matching system needs not only to follow personal preferences and appropriate dress codes, but also be creative when balancing factors such as color and style.

1.1 Background:
Fashion is a rapidly growing industry, which has motivated various research topics in the fashion domain, such as recommendation [3,4,5], retrieval [6,8], fashionability prediction [7,9,14], and clothing parsing [13,14,35], etc. In this section, we will provide a historical overview of the clothing matching system research. Although the research on clothing matching based on deep learning has become popular in recent years, many researchers have been involved in clothing matching in the early research. For example, as early as 2012, Si Liu et al. (2013) [12] developed a latent Support Vector Machine (SVM) based recommendation system, magic closet. Given the user-specified reference clothing (upper-body or lower-body) and a user-specified occasion, e.g., sporting or shopping, the magic closet system automatically recommends the most suitable clothing from online shops by considering the wearing properly and wearing aesthetically. Compared with clothing matching, the research of clothing recommendation is more extensive. Dandan Sha et al. (2016) [11] proposed a clothing recommendation method designed for shopping websites. In this method, a variety of features are extracted from the image and several attribute classifiers are trained to analyze the content of each feature under different attributes, so as to recommend clothing. Moreover, the research on personalized fashion recommendation has also aroused the interest of many researchers [14,15]. For example, Qiang Liu et al. (2017) [16] proposes a DeepStyle method for learning the style characteristics of items and preferences of users. Based on the learned style features and the BPR framework, personalized recommendation can be performed.

Deep learning enjoys a massive hype at the moment. The past few decades have witnessed the tremendous success of the deep learning (DL) in many application domains such as computer vision and speech recognition. The academia and industry have been in a race to apply deep learning to a wider range of applications due to its capability in solving many complex tasks while providing start-of-the-art results [17]. Recently, deep learning has been improving the performance of the clothing matching system dramatically. Recent advances in deep learning based clothing matching systems have gained significant attention over overcoming obstacles of conventional models and achieving high quality clothing matching. Deep learning is able to effectively capture various features of clothing, and enable the codification of more complex abstractions as data representations in the higher layers. Furthermore, it catches the intricate relationships within the data itself, from abundant accessible data sources such as contextual, textual and visual information [18]. Vignesh Jagadeesh et al. (2014) [19] applied a data-driven approach to fashion recommendations for the first time and confirmed that the power of data-driven analysis is over perceptually driven models. data-driven Models can learn intuitive domain insights directly from the data, without requiring human in the loop.

The development of deep learning has greatly increased the demand for clothing matching datasets and their labeling. But manual tagging is extremely difficult: on one hand, a fashion concept is often subtle and subjective, and it is nontrivial to get consensus from ordinary labels if they are not fashion experts. On the other hand, there may be a large number of attributes for describing fashion, for which it is very difficult to obtain exhaustive labels [20]. So, several pioneering researchers have turned to other sources, where rich data can be harvested automatically. For example, Hu et al. (2015) [21] investigated the problem of personalized outfit recommendation with a dataset collected from Polyvore. McAuley et al. (2015) [22] proposed a general framework to model the human visual preference for a given pair of objects based on the Amazon real-world co-purchase dataset.

1.2 Overview of Clothing Matching Systems and Deep Learning:
Before we dive into the details of this survey, we start with an introduction to the basic terminology and concepts regarding clothing matching system and deep learning techniques.

a) Clothing Matching Systems:
Clothing matching refers to coordinating complementary fashion items such as the tops and bottoms in styles and colors to achieve a decent and generous effect on the whole. Clothing matching is a new research branch. The earliest research on clothing collocation was done by Iwata et al. [23] in 2011, a topic model was proposed to recommend tops for bottoms with a small dataset collected from magazines.

In recent years, with the development of deep learning, more and more research efforts dedicated to the study of clothing matching, which can be roughly organized into two groups: collaborative methods [24,25] and content-based methods [26]. The former one recommends fashion items that people with similar tastes and preferences liked based on their historical behaviors [27], such as the users' purchase behaviors [28], users' textual descriptions [29] and the behaviors of other users [30]. The latter tackles the problem by modeling the human preferences between fashion items based on their visual similarity [31,32].

*

Copyright © 2020, IEASRJ. This open-access article is published under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License which permits Share (copy and redistribute the material in any medium or format) and Adapt (remix, transform, and build upon the material) under the Attribution-NonCommercial terms.
b) Deep Learning:
Convolutional Neural Network (CNN): A CNN is a multi-layer network having convolutional layers. Each layer consists of a number of filters. A filter is applied to the various local sub-regions of the image giving rise to a feature map. A CNN requires fewer parameters as the features are directly extracted from content such as product image or product review text. In a deep multi-layer CNN, the different layers learn different image features. A common technique is to extract features from an image of a product using a pre-trained CNN.

Recurrent Neural Network (RNN): RNNs model sequential data, capture memory of the past and can be trained by unfolding and doing back propagation through time. However, simple RNNs often fail to capture long distance temporal dependencies. Long Short Term Memory network (LSTM) and Gated Recurrent Unit (GRU) have been proposed to address this issue and have performed very well. They enable recommender systems to model content sequences or temporal dynamics. RNNs have been used for session-based recommendations. Chaoyuan Wu et al. (2017) uses LSTMs to capture temporal dependencies of users and items.

Graph Neural Networks (GNN): The research on GNNs is closely related to graph embedding or network embedding, another topic which attracts increasing attention from both the data mining and machine learning communities. GNNs are deep learning models aiming at addressing graph-related tasks in an end-to-end manner. Many GNNs explicitly extract high-level representations. The main distinction between GNNs and network embedding is that GNNs are a group of neural network models which are designed for various tasks while network embedding covers various kinds of methods targeting the same task. Therefore, GNNs can address the network embedding problem through a graph autoencoder framework.

2. METHODS: STATE-OF-THE-ART:
In recent years, numerous new clothing matching methods have been proposed, with a growing emphasis on neural network-based approaches. In this survey, we will focus on the study of clothing matching by using clothing content to learn clothing compatibility, as is most commonly studied. Another case is collaborative methods.

2.1 Content-based Clothing Matching:
In the fashion domain, discovering items that are functionally complementary or visually compatible is important. Previous work on the problem of fashion compatibility prediction uses models that mainly perform pair-wise comparisons between items based on item information such as image, category, title, etc. There are many methods, such as graph neural networks, Bayesian Personalized Ranking (BPR), image embedding strategy, etc. In the following chapter, we will introduce these clothing matching methods respectively.

a) Bayesian Personalized Ranking (BPR):
BPR was proposed to solve the ranking task between items. BPR is a ranking algorithm, which is an algorithm framework that can be applied to various existing recommendation algorithms. The modeling objective of this algorithm framework is to optimize the order between items and solve the ranking task between recommended items. Another modeling objective is to solve the problem that the implicit feedback data cannot show the user preference.

In 2017, Xuemeng Song et al. proposed a content-based neural scheme to model the compatibility between fashion items based on the Bayesian Personalized Ranking (BPR) framework. This can jointly model the coherent relationship between different modalities of fashion items and their implicit preference, as shown in Figure 1. Then, many researchers have proposed many new fashion clothes matching and recommendation scheme based on improving the method of Neural Compatibility Modeling. Such as, unlike the most basic convolutional neural network trained on ImageNet to extract visual features, Guangyu Gao et al. (2019) chose to represent such a similarity function for feature extraction in terms of a deep convolutional neural network. Meanwhile, Song et al. constructed the FashionVC Dataset, Guangyu Gao et al. constructed a more unsupervised dataset, named MN Fashion, from a huge amount of fashion items from MicroBlog.

Figure 1: Neural Compatibility Modeling. A dual autoencoder network was employed to learn the latent compatibility space, where they jointly model the coherent relationship between visual and contextual modalities and the implicit preference among items via the Bayesian Personalized Ranking. C: category, T: title. “~” indicates the category hierarchy.
Specifically, to solve the latter two problems, the author uses the latent attribute-based compatible prototypes learned by the above model as templates to identify the discordant attributes in incompatible outfits. Then, the semantic representation of the incompatible outfit is modified based on the found discordant attributes, and the new item is retrieved based on the modified semantic representation. Firstly, the author finds the most similar compatible prototype \( P \) with the incompatible outfit \( g \), by calculating the Euclidean distance, and calculates the attribute-wise distance of \( g \) and \( P \), so as to identify the most discordant attribute that causes the incompatibility as follows:

\[
d_e(i,k,l^*,z) = \frac{||d^k_j - P^k_l||_z}{M_z}
\]

\[
z^* = \arg\max_{z} d_e(i,k,l^*,z)
\]

Where \( d_e(i,k,l^*,z) \) and \( M_z \) denote the attribute difference between \((t,b)\) (top \( t \) and bottom \( b \)) and \( P \) regarding the \( z \)-th attribute and the number of possible values for the \( z \)-th attribute, respectively. Based on the semantic attribute representation of the new outfit, the items that can be replaced can be retrieved by Euclidian distance, thus the task of clothing matching is fully completed.

To the same end, Xun Yang et al. (2019) proposed a solution named Attribute-based Interpretable Compatibility (AIC) method, and constructed a Lookastic dataset with fashion attributes available. Similar to \cite{38}, given a corpus of matching pairs, they first learn the interpretable matching patterns that lead to good matches. Specifically, the author first automatically extracted decision rules on matching prediction by using a decision tree method. Then, they designed an embedding module to explicitly learn the vector representation for each rule by preserving the semantics of attributes in the rule. They further proposed a joint modeling module that unifies the visual embedding and attribute-based rule embedding to predict the matching score. To enhance the interpretability, they designed an attention network to select the most informative matching patterns, making the overall prediction process easy-to-interpret.

\( b) \) Graph Neural Networks:

Extending neural networks to work with graph structured data was first proposed by Gori et al. and Scarselli et al. The interest in this topic resurfaced recently, with the proposal of spectral graph neural networks and its improvements. In fact, graph neural networks have been applied to product recommendation, which is similar to product compatibility prediction\cite{42}.

Recently, Guillem Cucurull et al. (2019) address the compatibility prediction problem using a graph neural network that learns to generate product embeddings conditioned on their context. Specifically, the author used the graph auto-encoder framework, which defines an encoder that computes node embeddings and a decoder that is applied on the embedding of each product. They represented clothing items and their pairwise compatibility as a graph, where vertices are the fashion items and edges connect pairs of items that are compatible; they then used a graph neural network based model to learn to predict edges. The network structure is as shown in Figure 3.
3. CONCLUSIONS:
In this article, we provided an extensive review of the historical research of clothing matching system and the most notable works to date on deep learning based clothing matching systems. We introduced related research work according to the classification of clothing matching research methods, and highlighted a bunch of influential research methods. Both deep learning and recommender systems are ongoing hot research topics in the recent decades. We hope this survey will enable readers to have a deeper understanding of this field, and hope more and more good works will emerge.

REFERENCES:
XV. Iwata, T., Watanabe, S., & Sawada, H. (2011). Fashion coordinates recommender system using photographs from This image, on conference on artificial intelligence.
ing systems.


