Member name: Louisa Shi/罗雨诗

High school: The High School Attached to Tsinghua University

Province: Beijing

Country/region: China

Instructor: Jia Jia

Title: Clothing Aesthetics Modeling Based on Deep Learning

本参赛团队声明所提交的论文是在指导老师指导下进行的研究工作和取得的研究成果。尽本团队所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果。若有不实之处，本人愿意承担一切相关责任。

参赛队员：罗雨诗　　指导老师：

2017年9月15日

Title: Clothing Aesthetics Modeling Based on Deep Learning

Abstract: Clothing is one of the symbol of modern fashion trends, research concerning clothing detail' s recognition still remains a hot topic in the field of computer. Existing research is mainly focused on relating clothing detail features and aesthetic styles, realizing the possibilities of auto clothing aesthetic appreciation. In this paper, based on existing data sets, we will introduce the ideology of Deep Learning into the association model between clothing visual features to improve the model' s effectiveness and accuracy. By calculating the distance between aesthetic words and seed words, we are able to map the most often used aesthetic words in a two-dimensional space, building a fashion semantic space. Later, we introduce the ideology of Deep Learning into the association model between clothing detail features, and designed experiments to prove the effectiveness of the model. In the research, we discussed how different parameters and structures in Deep Learning effect the outcome of the relating model, and by comparison we choose the best parameters of the model. Based on the work mentioned above, we analyze different styles of different categories' clothing in shopping websites, which further confirms the effectiveness of our model.

Keywords: clothing aesthetic appreciation, fashion semantic space, deep learning

Contents

1 Introduction

With the rapid development of technology and economic in daily lives, people' s standard of living has reached a new level, with more and more people starting to pay attention to their appearance. As an essential part of one' s appearance, clothing is highly valued by many people. Clothing can reflect one' s inner spirit, and also represent one' s self cultivation and personality. Choosing proper apparel has become a routine in our everyday life. Yet when faced with various clothes, picking out a suitable style is inefficient by using only observation and bare hands. Therefore, realizing automatic style recognition is a new approach to the problem. Enabling computers to understand aesthetics and admiring beauty have become a popular topic, which are also the goal we hope to achieve in this paper. Yet it is a difficult task to make it possible for computers to understand the meaning of beauty. Commonly, computers are programmed to simulate logic thinking of the brain through programming and algorithms, while human understanding of aesthetics lies within our perceptual cognition, making it hard to calculate. Studying how to make computers learn to appreciate beauty is a meaningful task. If given the ability to appreciate beauty, it would greatly reduce the inconvenience for masses to follow fashion trends and furthermore improve people's quality of life, also explore the possibilities of which artificial intelligence can achieve.

Aiming at clothing classification and parsing, a lot of researches has been done domestically and internationally. Wei Yang[1] built a application system which can identify details of apparels. Kota Yamaguchi[2] researched on how to distinguish similar styled clothes through comparison. Masaru Mizuochi[3] built a clothing searching system that can search with only pieces of information. However, researches mentioned above are all limited to the recognition of clothing details, without relating visual effects with the aesthetic style of clothing. Latest researches expanded the research field of computer automatic analysis and recognition of the aesthetic effects of clothing from the understanding of perceptual beauty. Jia Jia[4] connected visual features, Fashion Semantic Space and aesthetic vocabularies, building a framework which includes three layers, and bridging the gap between visual features and aesthetic words. Yihui Ma[5] collected and labeled a large dataset of clothing images, and proposed a model to evaluate the aesthetic styles of clothing. These studies mainly use autoencoder[1] in the modeling of clothing features aesthetic effects. In

[1]http://deeplearning.stanford.edu/wiki/index.php/Autoencoders_and_Sparsity

this paper, we will introduce the idea of deep learning in the modeling process based on a public dataset Yihui Ma[6] released, hoping to improve the accuracy of the automatic appreciation on clothing aesthetic effects.

In this paper, we introduce the idea of deep learning into the modeling between existing detailed clothing feature data and aesthetic styles. We use a public data base released by Yihui Ma[6], including 43596 professional fashion photos in ten years from vogue.com, and 60004 various photos from e-commerce website jd.com. Based on the dataset and the labeled features of clothing, we used Deep Neural Network (DNN) in association modeling between clothing features and aesthetic words, establishing a mapping from visual details to aesthetics. To prove the effectiveness of our model, we design several experiments to evaluate different parameters' mapping effects, including iterations, hidden layers and neurons number. By running comparative tests, we are able to choose the best set of parameter, and compare the result with the classic SVM model. The experiment shows that under 300 iterations, 6 hidden layers and 80 neurons per layer, the result of the DNN model can reduce it' s errors from 0.2083 to 0.1862, proving the effectiveness of the DNN model.

The contributions of this paper can be concluded as following:

- We introduce the ideology of deep learning into the modeling from clothing visual features to aesthetic styles, and improve the accuracy of the automatic appreciation of clothing aesthetics.
- We discuss the effects on the experiment results through adjusting parameters, then choose the best set of parameters. We improve the modeling effects by approximately 10% in terms of MSE compared to the SVM model.
- We present some interesting case studies to reveal the difference of clothing styles based on e-commerce data with our proposed model.

2 Related Work

In modern computing researches, there exist researches aiming at clothing recognition and classification, therefore we will introduce related works domestically and internationally under this topic.

2.1 Research on clothing recognition and classification

Wei Yang[1] constructed a comprehensive application system to analyze clothing pictures and generate accurate clothing labels. The model first syncopated the images, created different areas, and then correlated them with different elements to construct a clothing analytical model. Research shows that the system can segment images and identify clothes accurately. Kota Yamaguchi[2] selected similar styles from specified clothing images, using a large scale of images to train model to learn and identify the details of clothing. Experimental results show that this method can improve the accuracy of clothing details' labeling and parsing. Masaru Mizuochi[3] built a searching system which can retrieve clothes through partial details. By extracting the relations between parts and whole, the system realized retrieving clothing styles according to details. The study increased users satisfactory by nearly 20 percent. Kevin Lin[7] built a layered deep searching framework. The first layer is the visual expression learning of pre-training network model. In the second layer he increased the numbers of hidden layer to learn hash type, using it to achieve higher retrieval speed. The retrieval framework performed well when searching for large-sized clothing images.

All studies mentioned above have achieved good results, and can extract clothing features from clothing images accurately. These studies provide us with technical support for further researches on clothing recognition. However, these works are limited to the recognition of clothing details, and do not relate the visual details to the aesthetic styles of clothing.

2.2 Research on aesthetic styles based on clothing visual features

In order to link visual details of clothing with aesthetic styles, some studies have made deeper investigations related to this connection. Jia Jia[4] introduced a middle layer between clothing visual features and aesthetic words, defined a three-layer framework: visual feature - fashion semantic space - aesthetic lexical space, and created a connection between visual features and aesthetic words. The experiment shows that this three-layer framework can better establish the relationship between the clothing features and the aesthetic words. Yihui Ma[5] built a large dataset of clothing images, and constructed an automatic matching model which was able to analyze and evaluate clothing collocations. By exploring the hidden rules between

tops and bottoms, the model made it easier to appreciate clothing aesthetics. Yejun Liu[8] designed an interactive aesthetic evaluation application system named Magic Mirror, associating aesthetic words with visual characteristics which enables the computer to appreciate aesthetic styles automatically. The system can be used to evaluate various aspects of clothing based on fashion trends. The research showed that the system is accurate in judging the style of clothing. Jingtian Fu[9] further modified and improved the Magic Mirror based on Yejun Liu's study[8], meeting more personal needs. The system increased personal preferences to the fashion styles, realizing personal recommendations based on the genetic algorithm.

The researches mentioned above expanded the research field of computer' s automatic analysis and recognition of the aesthetic effects of clothing. In this paper, we will draw upon the data and public resources mentioned above. But these works mainly used auto-encoders for modeling the relations between clothing features and aesthetic effects, not applying the DNN model in the scenario. In this paper, based on Yihui Ma' s dataset[6], we apply the DNN model to map the clothing visual features to the aesthetic styles, striving to achieve a better accuracy on computers' automatic appreciation of clothing aesthetic effects.

3 Problem Definition

In this section, we will give a clear definition of the modeling between clothing features and aesthetic effects.

3.1 The definition of the fashion semantic space

In this paper, we will refer to the research work of Jia Jia[4] which uses the two-dimensional aesthetic coordinates based on Shigenobu Kobayashi[10]'s theory. The two-dimensional aesthetic space is a two-dimensional space with one dimension ranging from warm to cool and the other dimension from hard to soft. Words which can be used to describe clothing styles in daily life, such as casual, cool and so on, will be put into WordNet::Similarity[11] to calculate their distances from the seed words Shigenobu Kobayashi[10] defined. Then, the new words are into the two-dimensional aesthetic space.

3.2 Methods

We will carry out the research in three aspects:

- Exploring the latest international public dataset on clothing aesthetics [5].
- Establishing the aesthetic evaluation space using the similar way to define the two-dimensional aesthetics space mentioned in Jia Jia' s research[4].
- Based on the works above, introducing the DNN model into the modeling between clothing details and aesthetic words.

4 Dataset and features

4.1 An overall description of the dataset and examples

In order to model the relationship between the characteristics of clothing and aesthetic styles, we need a dataset of clothing images which is comprehensive and large-scale. Therefore, this article uses an international public clothing aesthetics standard dataset published by Yihui Ma[5].

Figure 1: Examples of the fashion dataset.

The first part of the dataset is downloaded from a professional fashion website vogue.com with a collection of 43596 fashion show pictures, including men's and women's clothing (nearly the same proportion). Most of the images includes top and bottom parts of an outfit at the same time. The second part of the dataset

includes a total of 60004 clothing images downloaded from the e-commerce website jd.com, in which men' s and women' s clothes are about the same proportion, and most of the pictures contain only single tops or bottoms. Some image examples are shown in Figure 1. The labeling details can be seen at Table 1.

	Feature	#		Feature	#
Top	Gender	3	Pants	Gender	3
	Length	5		Length	4
	Sleeve	5		Waist	3
	Collar	8		Pattern	9
	Opening	7		Material	9
	Model	5		Model	3
	Pattern	9		Type	9
	Material	9	Skirt	Length	3
	Type	15		Model	2
Others	Occasion	8		Pattern	9
	Season	3		Material	9
	Style	2		Fold	2

Table 1: Feature annotation details.

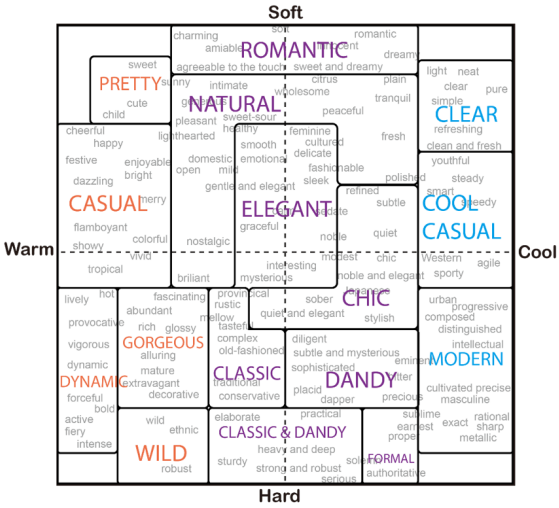

Figure 2: Fashion Semantic Space.

4.2 Detailed explanation of feature annotations

Specifically, the gender features of the top are divided into three categories: male, female and general. The feature of the length is divided into: short, medium, long, umbilical, super long. Sleeve length is divided into sling, sleeveless, short-sleeved, medium sleeve, long sleeve. The neckline is divided into a fur collar, a v-neck, a collar/a wide open collar, a round collar, a lapel, a turtleneck, and a hooded hat. The opening is divided into single-row buttoned, half-row buttoned, double-row buttoned, pullover, zipper, half row, zipper, front open. The dress type is divided into tight, straight tube, loose, cape, tight waist. Texture is divided into solid colors, plaids, dots, floral, horizontal stripes, vertical stripes, Numbers + letters, patterns set, pattern repetition. Materials are blended, knitted, silk/chiffon/yarn, leather, jeans, cotton, chemical fiber, flax, wool. Attribute of the clothes is divided into suits, sportswear, ski-wear, Down, shirts, sweaters, t-shirts, fleece, coats, leather coat, jacket, vest, fur, dress, cheongsam. Similarly, skirts and other type of cloth follow the same feature annotation. For more detailed features labeling please see appendix A. Specially, attributes for clothing category features, such as sweaters, shirts, etc., do not belong to the basic features of clothing. Style, namely clothing aesthetic effect, correspond to a point on the two dimension space, as shown in the

Figure 2. The study in Yihui Ma[6] mentioned that they recruit part-time annotators remotely made the annotations on the website. System gives them problems of the characteristics of clothing pictures, and they make choice that conforms to the option of annotation. In order to ensure the accuracy of the annotation results, each image characteristics is labeled three times by different people, and the final result is the majority of the three results.

According to the above classification and labeling method, Yihui Ma[6]'s team take on 43596 fashion show pictures from vogue.com in the past decade and 60004 clothing products pictures from jd.com containing different style. Some samples of the data are shown in Figure 3. Our research of modeling deep learning clothing aesthetic appreciation is based on this dataset.

Figure 3: Examples of the feature annotations.

5 Modeling of clothing aesthetic appreciation based on deep learning

In this chapter, we will introduce the application of deep learning model in this paper.

5.1 Introduction about DNN

DNN (Deep Neural Network), also called Deep Neural Network, is a model that can learn the mapping between input and output, containing several layers of neurons. Each layer will be calculated by the weight linear calculation and going through the activation function. The corresponding output vector can be obtained according to the specific input vector. The layers can be divided into three types: input layer, hidden layer and output layer, as shown in Figure 4[2]. Generally, the first layer is the input layer, the last layer is the output layer, and the rest of the middle layers are hidden layers.

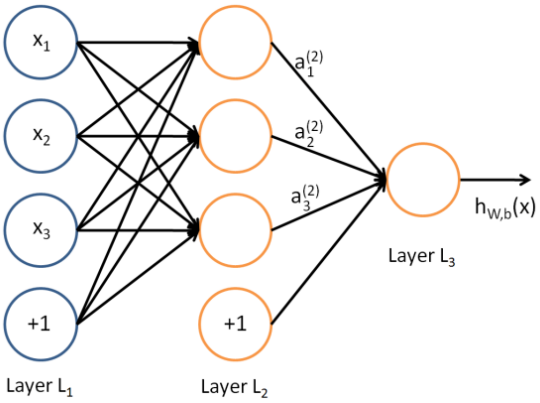

Figure 4: Basic structure of DNN.

5.2 Aesthetics-oriented modeling using DNN

In this section, we will introduce the method of using DNN to build the aesthetics-oriented mapping formally. In Table 2, we list the definitions of all the mathematical notations.

5.2.1 Problem Formulation

Given a set of clothing images V, for each image $v_i \in V$, we use an N dimensional vector x_i=$\langle x_{i1}, x_{i2}, \ldots, x_{iN} \rangle (\forall x_{ij} \in [0, 1])$ to indicate v_i's visual features. X is defined as a $|V| * N$ feature matrix with each element x_{ij} denoting the jth visual features of v_i.

Additionally, we use $Y(wc, hs)(\forall wc, hs \in [-1, +1])$ to represent the Fashion Semantic Space. The horizontal axis represents warmest to coolest with coordinate value wc varying from -1 to +1, while the vertical axis represents hard-soft with hs

[2]http://ufldl.stanford.edu/wiki/index.php/Neural_Networks

Symbol	Definition	Symbol	Definition
V	the image dataset	$h_i^{(l)}$	the lth layer's vector of v_i
v_i	the ith image in V	$W^{(l)}$	weight between lth and $l+1$th layers
x_i	the visual features of v_i	$b^{(l)}$	bias between lth and $l+1$th layers
N	the dimension of x_i	$\hat{wc}_i$	the predicted value of wc
Y	Fashion Semantic Space	J	cost function in model training
wc	warm-cool, 1st dimension in Y	m	the number of image samples
hs	hard-soft, 2nd dimension in Y	λ_1, λ_2	the hyperparameters in J
M	the mapping from (V, X) to Y	θ	the parameters in gradient descent
N_h	the hidden layer number of DNN	α	the step size in gradient descent

Table 2: The notations.

varying from -1 to +1. Focusing on fashion styles, a series of fashion descriptors (e.g. elegant, sporty, natural) that people usually use to describe the styles of clothing are coordinated in Y.

The problem we study on is to find the coordinates in Y for a given clothing image. According to the coordinates, we can get a few fashion aesthetic words with closest distances in Y, and regard them as the fashion style of the input image. Therefore, we need to train a mapping function$M : (V, X) \Rightarrow Y$ to accomplish the task of mapping clothing visual features to the Fashion Semantic Space.

5.2.2 Model Structure

For each dimension of $Y(wc, hs)$, we train a separate DNN model, and combine the mapping results as the coordinates in Y. In this way, we map the visual features to the Fashion Semantic Space successfully. Due to the similarity between the two DNNs, we take warm-cool dimension as an example and introduce the details of the DNN structure and its training process.

Supposing DNN has N_h hidden layers, the recursion formula between two adjacent layers is:

$$h_i^{(l+1)} = tanh(W^{(l)} h_i^{(l)} + b^{(l)}) \tag{1}$$

where $h_i^{(l)}$ denotes the lth layer's vector of v_i, $W^{(l)}$ and $b^{(l)}$are the weight and bias parameters between lth layer and $(l+1)$th layer , and $tanh$ is the tanh function $(tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}})$. Specially, $h_i^{(0)} = x_i$ and $\hat{wc}_i = h_i^{(N_{h+1})}$。

The cost function to evaluate the difference between the output layer values $\hat{wc}_i$

and the real values wc_i is defined as:

$$J(W,b) = \frac{\lambda_1}{2m}\sum_{i=1}^{m}||wc_i - \hat{wc}_i||^2 + \frac{\lambda_2}{2}\sum_{l}(||W^{(l)}||_F^2 + ||b^{(l)}||_2^2) \tag{2}$$

where m is the number of samples, λ_1, λ_2 are hyperparameters to control the relative importance of the two terms, and $||\cdot||_F$ denotes the Frobenius norm. The first term indicates average error between $\hat{wc}$ and wc. The second term is a weight decay term for decreasing the values of the weights and preventing overfitting.

For the training process of DNN, we define $\theta = (W, b)$ as our parameters to be determined. The training of DNN is to find a specific set of parameters θ^* which minimizes the cost function $J(W, b)$:

$$\theta^* = \arg\min_{\theta} J(W,b) \tag{3}$$

The optimization method we adopt is Stochastic Gradient Descent Algorithm [12]. For each iteration, we perform updates as following:

$$W = W - \alpha\frac{\delta}{\delta W}J(W,b) \tag{4}$$

$$b = b - \alpha\frac{\delta}{\delta b}J(W,b) \tag{5}$$

where α is the step size in gradient descent algorithm.

After training two DNNs for each dimension of Y as above, our model can utilize the final parameters to calculate wc and hs for arbitrary input visual features x. Combining them as $y(wc, hs)$, we can find the corresponding coordinates in Fashion Semantic Space, and get related aesthetic words finally.

6 Experiments

6.1 Dataset

In this section, we will compare several experimental results of the deep learning model and the ordinary machine learning regression algorithm, which can verify the accuracy and effectiveness of our deep learning model. We will carry out the experiments on the dataset we introduce in Section 4.

6.2 Experimental Setup

In this paper, we use DNN as the model of deep learning and SVM as the basic machine learning model. SVM (Support Vector Machine) is a very classic algorithm

which is widely adopted in solving classification and regression problems over the past few decades. Therefore, we choose SVM as the baseline of regression algorithm.

Our experiments are divided into two steps. First, we change the number of training iterations and the number of hidden layers and neurons respectively to see how they affect the DNN model results. Then, we compare the results with the SVM model result, as to prove the validity of our model.

6.2.1 The effects of different parameters and structures on DNN model results

- **The experimental results influenced by changing the number of iterations.** The structure of neuron network is fixed, and the number of iterations is adjusted. Each time the final result will be the average MSE computed using five-fold experiments' data. The number of iterations is increased by 50 each time, and six sets of data are obtained.

- **The effect of layer numbers on the experiment results.** The optimal iteration number above is used. The number of neurons is fixed. The number of neurons is adjusted. Each time the final result will be the average MSE computed using five-fold experiments' data. Each time one layer is added and six sets of data are obtained.

- **The effect of the number of neurons on the experiment results.** Number of the layers is fixed. The number of neurons is adjusted. For every set of parameters, three experiments are carried out. The final result will be the average value of the MSE for each group. Each time one layer is added, and altogether 20 sets of data are obtained.

6.2.2 Comparison between DNN and SVM

Based on the results of the above experiments, the experimental results of DNN and SVM are presented to show the comparison between the deep learning model and the classic shallow machine learning model. The programming language used to describe DNN and SVM models in this article is Python (open source code : TensorFlow[3] and scikit-learn[4] respectively).

[3]https://www.tensorflow.org/

[4]http://scikit-learn.org/stable/

6.3 Metric

The experimental results in this paper are all evaluated by Mean Squared Error (MSE) to calculate the errors of the model. Smaller MSE value proves that the model is more accurate. The results of the experiment are all performed on five-fold cross-validation. The calculation formula of MSE is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(observed_i - predicted_i)^2 \tag{6}$$

6.4 Results and Analysis

6.4.1 Parameters' effects on deep learning model

In this section, we will report the effects of different parameters and structures in DNN on the associated modeling results. The impact of iteration number on experimental results is shown in Figure 5. We have tried the number of iterations with 50, 100, 150 and so on. As can be seen from the figure, the change of iteration number from 50 to 100 times has great influence on MSE which is greatly decreased. From 100 to 150 times the MSE is only slightly reduced. The MSE, between 200 and 300 times, is not significantly different and levels off. From 50 to 300 times, the error was reduced by 3.80%, thus the number of iterations to 300 or so is a desirable ideal number. So we choose 300 times as the iteration number for follow-up experiments.

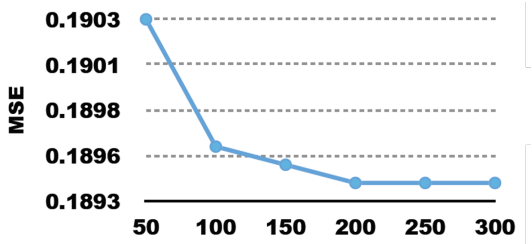

Figure 5: influence of the number of iterations on the correlation modeling results

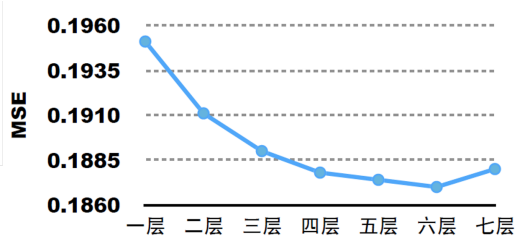

Figure 6: influence of the number of layers on the correlation modeling results

In Figure 6, we have tried one to seven layers respectively. The result shows that from one layer to seven layers, the MSE value reduced about 2.05%. From two to three layers, the MSE value was reduced by 1.09%. From three to four layers, it was reduced by 0.06%. From four to six layers, the change is relatively small. But in six to seven layers, the MSE grows instead, indicating that too many layers may make the error increase. So we choose six as the number of layers for the follow-up experiments.

Neuron number influence on the experimental results is shown in Table 3. We find that when the layer number is fixed, with the increase of neurons, MSE values gradually reduce between 4% and 8%. When the number of neurons is fixed, with the increase of layers, MSE values appear to drop after a rising trend. Comparing the experimental results of different groups, we select the number of iterations as 300, with a total of six hidden layers and 80 neurons in each layer, and get the best MSE result of 0.1862.

	10 units	20 units	40 units	80 units
2 layers	0.1954	0.1915	0.1897	0.1872
3 layers	0.1952	0.1923	0.1881	0.1877
4 layers	0.1930	0.1899	0.1886	0.1875
5 layers	0.1928	0.1900	0.1883	0.1875
6 layers	0.2040	0.1917	0.1893	0.1862

Table 3: The joint influence of layer size and neuron number.

6.4.2 The results of DNN and SVM

By changing the different parameters of the DNN model, the optimum value of the MSE in our experiments is 0.1862, and the worst experimental value we get is 0.204. On the contrary, the MSE of the SVM model is 0.2083. The best value for DNN is 10.6% lower than the errors from the SVM model. Thus, we concluded that through adjusting the parameters of the DNN model, we could get better results to establish a link between the characteristic of clothing and the aesthetic effects.

6.5 Case study

Using the above model, we have carried out some interesting case studies on the clothing styles in the e-commerce data. In particular, we took 100 pictures from the famous sports brand Nike' s online shopping website[5]. The pictures contain apparels of different ages and genders, different sports types, different seasons, and various types such as tops or bottoms. Each image is fed into our model and mapped to a point in our two-dimensional aesthetic space. From the results, we have some interesting observations.

- **Comparison among different genders and ages.** In Figure 7, the distribution of men's clothing is relatively concentrated, which is mostly in the

[5]https://nike.tmall.com/

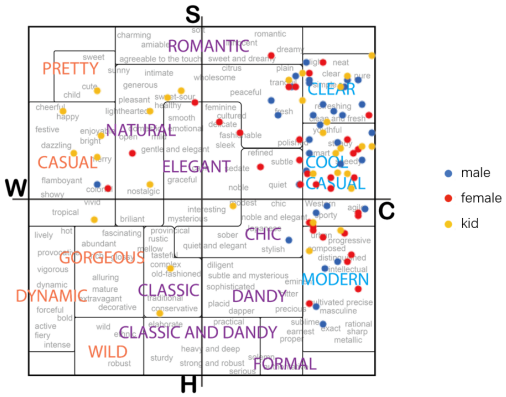

Figure 7: Different genders and ages.

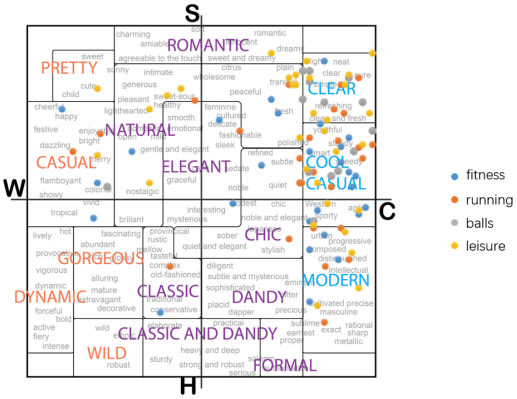

Figure 8: Different kinds of sports.

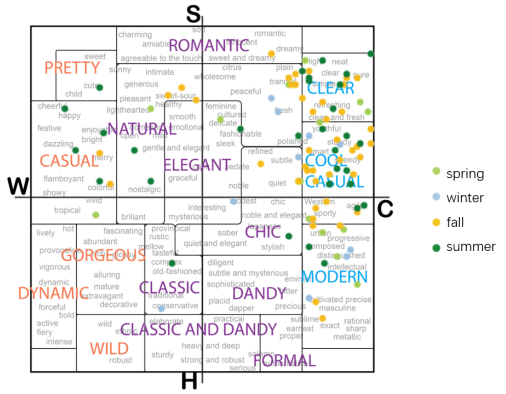

Figure 9: Different seasons.

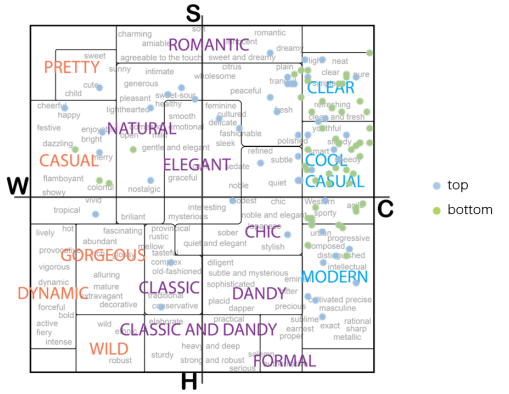

Figure 10: Tops and bottoms.

right position of the space, corresponding to clear, cool, casual styles and so on. However, the style of women's wear has been expanded slightly to the left. Besides the casual, modern styles, there are also some points in elegant, natural styles in the middle region. As for kids' clothing, it covers the entire plane, indicating that its style is very diverse.

- **Comparison among different sports.** In Figure 8, sports clothing for ball playing is more concentrated (centralized) distributed, and clothing for fitness, running, leisure is more dispersed. They are not limited to the far right area, and some even distribute into the left most area, suggesting that these categories of clothing contain some of the more unique designs.

- **Comparison among different seasons.** In Figure 9, there are some differences styles of clothing for different seasons. The spring and autumn outfits have very similar distributions. The distribution of winter clothing is relatively concentrated, while the summer clothing has a variety of styles because of its design diversity which covers the multiple regions in the space.

- **Comparison between tops and bottoms.** In Figure 10, we can see that tops are distributed a wider range than the bottoms, which means that the tops tend to have a more diverse design with rich colors, and the bottoms of sport brands (typically the Nike brand bottoms) is unitary. The result is consistent with our daily knowledge.

7 Conclusion

Clothing has always been a hot topic among people, and it is very meaningful to research upon clothing aesthetics. In this paper, we use the idea of deep learning to construct a model between clothing features and aesthetic words, intending to improve the effectiveness and accuracy of the associated model. First, we calculate the distance between aesthetic words and the seed words, map the aesthetic words on the two-dimensional aesthetic coordinates, and build a fashion semantic space. Then, we introduce the idea of deep learning into the modeling between clothing features and aesthetic words, and make comparison experiments to prove the effectiveness of the model. Lastly, we analyze the style differences on e-commerce dataset, further proving the efficiency of the model in the case-study.

References

[1] Wei Yang, Ping Luo, and Liang Lin. Clothing co-parsing by joint image segmentation and labeling. In *Computer Vision and Pattern Recognition*, pages 3182 – 3189, 2015.

[2] K Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg. Retrieving similar styles to parse clothing. *Pattern Analysis & Machine Intelligence IEEE Transactions on*, 37(5):1028–40, 2015.

[3] Masaru Mizuochi, Asako Kanezaki, and Tatsuya Harada. Clothing retrieval based on local similarity with multiple images. In *ACM International Conference on Multimedia*, pages 1165–1168, 2014.

[4] Jia Jia, Jie Huang, Guangyao Shen, Tao He, Zhiyuan Liu, Huanbo Luan, and Chao Yan. Learning to appreciate the aesthetic effects of clothing. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[5] Yihui Ma, Jia Jia, Suping Zhou, Jingtian Fu, Yejun Liu, and Zijian Tong. Towards better understanding the clothing fashion styles: A multimodal deep learning approach. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[6] 马依慧. 基于美学原理的服装搭配研究与系统构建. Bachelor thesis, 清华大学, 2016.

[7] Kevin Lin, Huei Fang Yang, Jen Hao Hsiao, Jen Hao Hsiao, and Chu Song Chen. Rapid clothing retrieval via deep learning of binary codes and hierarchical search. In *ACM on International Conference on Multimedia Retrieval*, pages 499–502, 2015.

[8] Yejun Liu, Jia Jia, Jingtian Fu, Yihui Ma, Zijian Tong, and Zijian Tong. Magic mirror: A virtual fashion consultant. In *ACM on Multimedia Conference*, pages 680–683, 2016.

[9] Jingtian Fu, Yejun Liu, Jia Jia, Yihui Ma, Fanhang Meng, and Huan Huang. A virtual personal fashion consultant: Learning from the personal preference of fashion. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[10] Shigenobu Kobayashi. The aim and method of the color image scale. *Color Research & Application*, 6(2):93–107, 1981.

[11] Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. Wordnet: similarity - measuring the relatedness of concepts. In *National Conference on Artifical Intelligence*, pages 1024–1025, 2004.

[12] L é on Bottou. *Large-Scale Machine Learning with Stochastic Gradient Descent*. Physica-Verlag HD, 2010.

A Appendix

TOP	
Gender	Female / Male / General
Length	Extra-Short / Short / Mid / Long / Extra-Long
Sleeve Length	Suspender / Sleeveless / Short / Mid / Long
Collar Shape	Round / Lapel / High / Stand / V / Bateau / Fur / Hoodie
Opening	Single-breasted / Double-breasted / Half-breasted / Zipper/ Half-zipper / Pullover / Open
Model	Tight / Straight / Loose / Waisted / Cloak
Pattern	Pure / Grid / Dot / Floral / Vertical stripe / Cross stripe / Number&Letter / Focus / Repeat
Material	Cotton / Chemical fiber / Blending / Woolen / Silk / Denim / Leather / Flax / Knit
Type	Dress / T-shirt / Sweater / Shirt / Suit / Jacket / Vest / Hoodie / Coat / Sportswear / Down-Jacket / Fur / Leather / Cheongsam / Mountaineering jacket

BOTTOM	PANTS
Gender	Female / Male / General
Length	Short / Mid / Long / Cropped
Waist	Low / Normal / High
Model	Tight / Straight / Loose
Pattern	Pure / Grid / Dot / Floral / Vertical stripe / Cross stripe / Number&Letter / Focus / Repeat
Material	Cotton / Chemical fiber / Blending / Woolen / Silk / Denim / Leather / Flax / Knit
Type	Leggings / Sport pants / Hot pants / Harem pants / Bell-bottoms / Suspender / Jeans / Suit pants / Casual pants

BOTTOM	SKIRT
Length	Short / Mid / Long
Fold	With / Without
Model	A-shape / Packet Hip
Pattern	Pure / Grid / Dot / Floral / Vertical stripe / Cross stripe / Number&Letter / Focus / Repeat
Material	Cotton / Chemical fiber / Blending / Woolen / Silk / Denim / Leather / Flax / Knit

Figure 11: The full annotated features of the dataset.