https://doi.org/10.1016/j.jvcir.2019.03.003

Fashion Recommendations through Cross-media Information Retrieval

Wei Zhou[a], P.Y. Mok[b,c,*], Yanghong Zhou[b,c], Yangping Zhou[b,c], Jialie Shen[d], Qiang Qu[a], K. P. Chau[c]

[a]*Shenzhen Institutes of Advanced Technology, CAS, Shenzhen, China*
[b]*The Hong Kong Polytechnic University Shenzhen Research Institute, China*
[c]*Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hong Kong*
[d]*Northumbria University, Newcastle NE2 1XE, UK*

Abstract

Fashion recommendation has attracted much attention given its ready applications to e-commerce. Traditional methods usually recommend clothing products to users on the basis of their textual descriptions. Product images, although covering a large resource of information, are often ignored in the recommendation processes. In this study, we propose a novel fashion product recommendation method based on both text and image mining techniques. Our model facilitates two kinds of fashion recommendation, namely, similar product and mix-and-match, by leveraging text-based product attributes and image features. To suggest similar products, we construct a new similarity measure to compare the image colour and texture descriptors. For mix-and-match recommendation, we firstly adopt convolutional neural network (CNN) to classify fine-grained clothing categories and fine-grained clothing attributes from product images. Algorithm is developed to make mix-and-match recommendations by integrating the image extracted categories and attributes information are with text-based product attributes. Our comprehensive experimental work on a real-life online dataset has demonstrated the effectiveness of the proposed method.

Keywords: Fashion recommendations, image retrieval, human parsing, image features

*Corresponding author
Email address: tracy.mok@polyu.edu.hk (P.Y. Mok)

1. Introduction

Total online retail sales continue to grow rapidly worldwide. Online sales volume has reached 414 billion dollars in the United States in 2018 [1, 2]. Among these online retail sales, fashion products rank number one in all categories. In view of the prevalence of online shopping, the value of product recommendation is increasingly recongised because it helps consumers effectively screen huge amount of data available online and identify the right products that meet their needs.

In contrast to traditional product recommendations, fashion recommendations have certain unique characteristics. Firstly, fashion recommendation is time sensitive. Fashion comes and goes that a customer selected a product last year but he/she may no longer prefer it anymore this year. Secondly, product information is often presented in different media, for example, fashion products are displayed online through textual description, images and videos. Images and videos are non-structural data that are difficult to use in recommendation without pre-processing. Thirdly, fashion recommendations must take into consideration several types of information, such as body shape and size of the users, which are usually not available when online recommendations are suggested. Lastly, in addition to similar product suggestions, mix-and-match should also be considered in fashion recommendation [3]. For example, if a user selects a jacket, pants, jeans and/or shoes that coordinate well with the selected jacket should be recommended to the user as mix-and-match recommendations.

Traditional fashion recommendations were done using expert rules based mainly on textural information of the fashion products. Image feature extraction [4, 5, 6] can extract hidden information in clothing images, thereby supplementing clothing textual description. Until recently, new technologies and methods [7, 8, 9] were proposed for clothing product retrieval, mix-and-match clothing item suggestion, and user-based personalised recommendation using image data. By extracting features from the images of fashion products, the

similarity between clothing items can be calculated by evaluating the similarity of different extracted image features and summarising all features through a weighted average score [10, 11]. For instance, researchers extracted various features from product images, starting from measuring image similarity, then sorting and indexing and finally retrieving products from these image contents [12, 13, 14]. Recommending similar clothing items has become a problem of feature extraction in image-based approaches. Previous clothing retrieval methods [15, 16] exploited the common characteristics of an image, such as colour, texture and shape features. For example, Hou *et al.* [17] used invariant moments and Fourier descriptors to identify the shape of a garment and combined the colour histogram to analyse clothing product image retrieval. Wang *et al.* [18] proposed a content based image retrieval (CBIR) approach on the basis of bag-of-visual-words model, in which a codebook was constructed from an extracted dominant colour palette. Song et al [19, 20, 21, 22] proposed a content-based neural scheme to model the compatibility between fashion items based on the Bayesian personalized ranking (BPR) framework. The scheme is able to jointly model the coherent relation between modalities of items and their implicit matching preference. Lie et al. [23, 24] proposed method which is able to learn the attribute-specific and attribute-sharing features via graph-guided fused lasso penalty.

Traditional image processing technology may be significantly 'shallow' to extract hidden information from fashion product images. An increasingly deep information can be recovered at present with the recent success of deep learning technologies [25, 26, 27]. Deep convolutional neural networks (CNN) can extract deep features from fashion images for various applications, including clothing style recognition and retrieval [28, 29], clothing recognition and retrieval [30, 3] and automatic product recommendation [31, 32]. We propose a fashion product recommendation method by cross validating extracted textual and product image information. Human parsing is the most crucial technique among many deep learning techniques in our proposal. Human parsing segments a human image into semantic fashion/body regions, from which further analysis can be

conducted on the basis of the segmented/parsed clothing regions. The key contributions are summarized as follows: Firstly, a re-categorisation method of fashion products is proposed to classify products from different online sources into the defined categories through text mining and semantic analysis. Secondly, deep convolution neural networks are trained to classify image types, product categories and attributes, and more importantly parse input human images into regions with semantic meaning. Thirdly, new methods are proposed to extract colour and texture features from the segmented product regions, and a novel Pantone-based colour similarity descriptor is developed. Lastly, similar and mix-and-match recommendation techniques are proposed based on both textual and human parsing features.

2. The proposed method

2.1. Method overview

We propose a fashion product recommendation method by cross validating extracted textual and product image information, as shown in Figure 1.

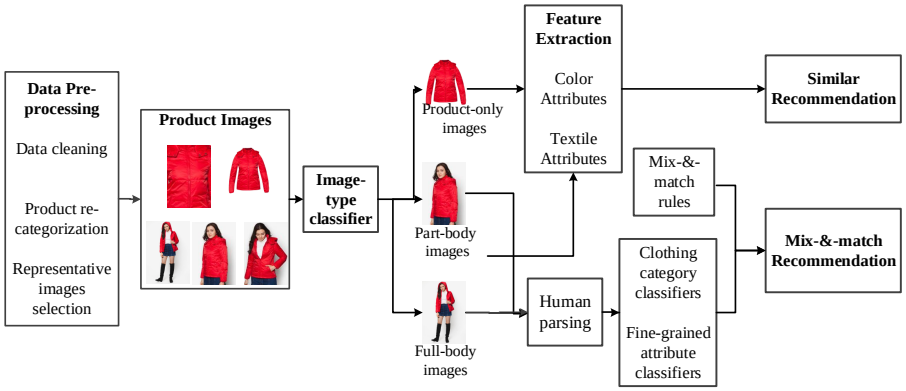

Figure 1: **Method Overview.**

Firstly, we develop a fashion keyword library through statistical natural language processing and then formulate algorithms to re-categorise fashion prod-

ucts into our defined taxonomy and automatically label fashion product attributes by text mining and semantic analysis. Secondly, we construct a CNN classifier for image mining to sort the product image types and identify the representative image for each product. Thirdly, we develop models to extract image contents, such as Pantone-based dominant colour and texture features, from representative images. Finally, we develop novel models by integrating text-based product attributes and image-extracted features to recommend similar products and mix-and-match items for the specific products. For product recommendation, we propose new similarity measures based on the image-extracted colour and texture descriptors to suggest similar products. For mix-and-match recommendation, we train deep CNN classifiers to obtain fine-grained clothing category from product images. We develop mix-and-match algorithms that suggest matching items using human parsing techniques and expert rules.

2.2. Data collection

Data used in this work were crawled from different online fashion websites, including Zalora(https://zh.zalora.com.hk), Uniqlo(https://www.uniqlo.com.hk/), H&M(https://www2.hm.com) and ASOS(https://www.asos.com). A total of 158,211 products were crawled from these websites. Among these products, 91,491 are products for women, and 66,720 are products for men, including 626,516 product images. The main features that we obtained from the online websites include product category, brand, gender, price, description and images. Two-types of data pre-processing are carried out. Firstly, clothing products are mapped into the defined categories. Secondly, a representative image is selected for each clothing product by using a trained network.

2.3. Text-based information processing

Product information is collected from different websites, and the product classification varies among these websites. To recommend products across websites, a new taxonomy of clothing products is defined to unify product categories. We propose a product re-categorisation method to map all products from the

original categories to our new redefined categories. We define a new taxonomy structure which consists of 101 product categories, with 42 categories for men and 59 for women. For each product category, a keyword library is defined. We assign a product to a category if the text description of the product hits any of the defined keywords (i.e. different text vocabularies expressing the same product category).

2.4. Image-based feature extraction

Text descriptions of a product on different websites may be inconsistent or incomplete. The text-mining technique described in the previous section can handle the issue of inconsistency. To handle incomplete description, we propose to extract information from product images, including clothing category and fine-grained attributes. Figure 1 illustrates the image-based processing steps in the proposed method. Specifically, it is crucial to distinguish the type of product images, such as full-body, half-body, clothing detail, product-only and other images that are irrelevant to the product itself. We construct a large image-type dataset for clothing products, consisting of 24,255 full-body, 52,249 half-body, 53,493 product-only and 28,848 product detail images. We develop the image-type classifier using VGG-16 net [33] on Caffe library, only change the channels of the final fully-connected layer to the number of classes. We train the image type classifier with defined image-type dataset. The dataset was split into three sets: 70% for training, 10% for validation, and 20% for testing. We train the network also by fine-tuning weights pre-trained on ImageNet dataset with a mini-batch stochastic gradient descent with a momentum of 0.9, and weight decay of 0.0005. The accuracy of the model on the validation dataset is 96.25%. The trained classifier will classify the type of each input human image and the associated image-type score.

We use the image-type classifier to select one representative image from the product images. The image can be used for the subsequent image processing for product feature extraction (e.g. colour and texture). The method of feature extraction may vary in accordance with the image types. If a product-only

image is presented, then we typically select it as the representative image. This type of image is commonly clean, with less occlusion and added information on the product itself being shown in the middle of the image. Therefore, it is advantageous in extracting dominant clothing colours and recognising fine-grained clothing category and attributes. If product-only image is not available, then we use the image with the highest score among the full-body images as the representative image. Otherwise, the half-body image with the highest score will be selected as the representative image. Product features are then extracted from the representative image. For product-only images, we can remove the image background by setting the RGB values for colour threshold masking. For full- and half-body images, human parsing is applied to segment the region of specific clothing for further analysis.

We adopted the part-detection based and CRFs embedded deep neural network developed previously [34] for human parsing. We train the network on the ATR dataset [35], which contains 7,702 images each is paired with a ground-truth. The dataset is split into two sets: the first set with 90% images for training and the second set with 10% images for testing. The network was trained by phase achieving over 93% pixel accuracy. Figure 2 shows an example of our human parsing model. Each image is segmented into 18 parts, including: Background, Hat, Hair, Sunglasses, Skirt, Upper-clothing, Pants, Dress, Left-shoe, Belt, Right- shoe, Face, Right-leg, Left-leg, Left-arm, Bag, Right-arm, Scarf. All the labels are not necessarily appear in an input image according to the character of the input image.

2.4.1. Dominant color histogram extraction

In this section, a dominant colour histogram extraction method based on Pantone colour is introduced. Pantone, Inc. is the world-renowned authority that publishes colour standards and facilitates selection and accurate communication on colour across specific industries. A total of 2,310 Pantone colours are selected [36] to cover most colours used in textile and fashion products. To compare product colours, we group the RGB values of all image pixels into a small

Input image Human parsing result using our method

Figure 2: **Human parsing example.**

number k of colour histograms through clustering technique. For each Image I, the dominant colour histogram can be represented using a list of two-tuples as follows:

$$H\left(I\right) = \left\{\left(rgb_1, per_1\right), \left(rgb_2, per_2\right) \cdots \left(rgb_k, per_k\right)\right\} \tag{1}$$

where $H\left(I\right)$ is the colour histogram extracted, and $\left(rgb_k, per_k\right)$ denote the RGB value of the k-th colour histogram centroid rgb_i and the percentage per_i it occupies in the image. The colour feature of each image can then be represented as a descriptor of Pantone colours. The original colour histogram is mapped to the nearest Pantone colour p_j $(1 \leq j \leq 2,310)$; thus, a new colour feature descriptor $P(I)$ is obtained for Image I.

$$P\left(I\right) = \left\{\left(p_1, per_1\right), \left(p_2, per_2\right) \cdots \left(p_k, per_k\right)\right\} \tag{2}$$

2.4.2. Textile texture of clothing products

The textile feature of clothing products are extracted from clothing images. HOG and LBP are typical features used to describe texture. However, we

construct a textile-based classifier. We construct a dataset and train a CNN textile-group classifier by defining 14 textile groups. In addition, we further extract feature in 47-dimension vector on the basis of the work of [37]. Generally, two kinds of textile features are extracted from the product region of the representative images.

2.5. Fashion recommendations

2.5.1. Similar product recommendation

The feature extraction-based CBIR method and human parsing-based dominant colour similarity are proposed. As discussed previously, text- and image-based product features are extracted. We can summarise all product features as follows:

$$C = \{f_1, f_2, f_3 \cdots f_n\} \tag{3}$$

where n is the number of features, and f_k is the k-th feature of the feature list. The similarity score of two products is calculated using Equation (4).

$$Sim\left(C_i, C_j\right) = \frac{\sum_{k=1}^{n} w_k Sim\left(C_{f_{k,i}}, C_{f_{k,j}}\right)}{\sum_{k=1}^{n} w_k} \tag{4}$$

where C_i and C_j are the two feature vectors of products i and j respectively; w_k is the weight for the k-th features of vector C, and $Sim\left(C_{f_{k,i}}, C_{f_{k,j}}\right)$ is the similarity score of the k-th feature. As shown, similarity computation mainly includes computing feature similarity and defining weight coefficient. In our similar product recommendation, three types of feature similarity, namely, continuous features, discontinuous features and Boolean features are computed.

We consider colour, textile, attribute and price features in similarity computation. For colour similarity, the similarity between Pantone colours must be defined, although we simplify the colour space through Pantone descriptor by Equation (2). We further group 2,310 Pantone colours into a small number m

of colour groups as manually defined by fashion design experts.

$$G = \{G_1, G_2 \cdots G_m\} \tag{5}$$

The guiding principle of colour group definition is that colours within the same group are all similar but different among various groups in human vision. Therefore, in this study, only similar colours within the same group are calculated. The pseudocode of colour similarity among colour histograms is presented in Algorithm 1.

Algorithm 1 Color similarity calculating between color histograms

Input: Two transformed color histograms $P(I_1)$ and $P(I_2)$; Grouped Pantone colors G.

Output: The similarity of two dominant color histograms s.

1: For elements in $P(I_1)$, $P(I_2)$, find distinct pantone set G'.
2: Merge elements if pantone centers are the same.
3: Normalization $P(I_1)$ and $p(y)$, making sure sum of all percentage $\sum_1^k per_k = 1$.
4: For each centroid e in G':
5: If e in both $P(I_1)$ and $P(I_2)$: $d = d' * abs(per_1 - per_2)$. d' is the RGB distance of two colors.
6: Else e not in both $P(I_1)$ and $P(I_2)$: $d = abs(per_1 - per_2)$
7: Get the total distance $D = \sum d$ in G_P.
8: Similarity of $P(I_1)$ and $P(I_2)$, $s = \frac{(2-D)}{2}$.

For clothing products, textile features that extracted from image data, as defined in Section 2.4.2, are continuous features. We compute the textile similarity using cosine similarity.

In addition to colour and textile features, both product attributes and price features are formulated as vectors of 0 and 1. For product attribute, the Boolean value indicate the absence and presence of the corresponding product attributes; for price groups, the value indicates different and similar in normalised price

Figure 3: **Examples of similar (first 3 rows) and mix-and-match (last 3 rows) recommendations**.

groups. The similarity of Boolean values can be calculated using (6).

$$Sim\left(f_i, f_j\right) = \begin{cases} 1, f_i = f_j \\ 0, f_i \neq f_j \end{cases} \tag{6}$$

2.5.2. Mix-and-match product recommendation

In addition to similar product recommendation, customers are often suggested with other fashion items that match well with a given item, which is called mix-and-match recommendation. Mix-and-match recommendations are traditionally implemented on the basis of expert knowledge.

In this study, we train residual networks [37] to recognise fine-grained cloth-

ing categories from images. In addition, we can use the trained human parsing network to segment all clothing regions from a full-body image, as illustrated in Figure 2. By doing so, not only the target product but also its matching items can be identified from full-body images. The clothing regions of matching items can be obtained, from which we recognise the corresponding product categories. We use the defined method for recommending similar products to identify, from re-categorised product dataset, the most similar products for each cropped matching item. With human parsing technique, we can use the defined similar product recommendation method to recommend mix-and-match items integrating both textual and image features, by defining a mix-and-match fitness score similar to that of Equation (2). Moreover, for each product category, expert rules are used to define matching categories and product attributes, such as colours. When full-body images are not available, we identify the matching categories based on expert rules, then identify matching items from each corresponding category with highest fitness score with the target product.

Figure 3 shows some examples of similar and mix-and-match recommendations suggested by our algorithms. There are 6 rows of products and its similar recommendations and mix-match recommendations with the first item as the target product. Rows 1 to 3 are for similar recommendations while rows 4 to 6 are for mix-match recommendations, which show representative images of suggested recommendations.

In similar recommendations, similar products are retrieved from the same product category. Similarity descriptors are considered such as colour histogram, product category and style. For example, hats with black belt are ranked top of the recommendation list. In mix-and-match recommendations, mix-and-match products are selected and ranked from different product categories. Mix-and-match rules, which is compiled statistics from fashion product images. For example, in suit product mix-and-match, pants, ties, belts, shoes and wallets are selected and ranked by the mix-and-match relevancy.

3. Results and experimental evaluations

3.1. Experiment setup

In this section, the proposed method is evaluated through an experiment over a subset of our dataset, covering some selected categories, as shown in Table 1. Among the selected categories, only data from ASOS website are used in the experiment. In the literature, manual rating is typically the method used to evaluate the effectiveness of recommendations, in which subjects are asked to rate the appropriateness of each recommendation with a 1-to-5 or 1-to-7 Likert scale. Assuming a product has five similar and five mix-and-match recommendations, with manual rating method, each subject must rate each of the ten recommendation against the given product individually. Considering the large number of product categories and large number of product in the testing dataset, such manual rating is obvious too tedious to implement and not recommended as experimental elevation in our study.

Table 1: Number of candidate products under different categories or wish-lists

Category/ Wish-list name	**Number of products in the category**
Dress (F)	22864
Top (F)	1533
Skirt (F)	2284
Jacket (F)	389
Cardigan (F)	342
Shoes (F)	11620
Bag (F)	5978
Shirt (M)	2165
Suit (M)	289
Pants & Trousers (M)	818
T-shirt (M)	6829
Jeans (M)	1381
Shoes (M)	5235
Bag (M)	3246

To avoid human fatigue, we designed a new experiment method by preparing pseudo recommendation lists. Subjects were asked to randomly select product from our testing dataset using a mobile device. Each selected product is presented with a pseudo-recommendation list of 20 similar and 20 mix-and-match products, similar to usual practice in online shopping.

From the pseudo recommendation lists, subjects were asked to select what they consider the most similar or the best match with the given products, and add to specific wish-list. Different wish-lists were created for each product categories. In the entire process to assess product recommendations, subjects are shown with all images and textual information, but not the sources of recommendation. The accompanied video of the paper was shown to subjects to illustrate the experiment requirement.

In similar product recommendations, each pseudo recommendation list cov-

ers products from three sources, namely,

(i) candidates generated from the recommendation algorithms with highest scores,

(ii) candidates recommended in the original website of ASOS and

(iii) random noises.

We can evaluate (1) whether the proposed recommendation algorithms are comparable to human experts by comparing the sources of the products being selected by users/subjects because ASOS appoints uses human experts to recommend styles and (2) whether the proposed recommendation algorithms are better than a random walk.

A total of 95 undergraduate and postgraduate students, including 57 females and 38 males, from The Hong Kong Polytechnic University participated in our experiment from 6 to 18 February 2018. Products being presented to subjects are gender specific. In other words, female subjects only evaluate fashion recommendations for female products, and male subjects only evaluate recommendations for male products. A total of 655 wish-lists were recorded (i.e. 655 products), from which 132 wish-lists were removed because of data cleaning. We then analysed the remaining 523 records obtained from 95 subjects (users).

3.2. Similar recommendation results

Table 2: Comparison of expected numbers and actual numbers being selected from each source in similar product recommendations.

	Source a_1: by algorithm			Source a_2: by website/experts			Source a_3: random walk		
	(I)	(II)	(III)	(I)	(II)	(III)	(I)	(II)	(III)
all	44%	42%	14%	47%	46%	7%	74%	15%	11%
Cardigan (F)	44%	31%	25%	38 %	62%	0%	85%	15%	0%
Dress (F)	59%	20%	20%	47%	53%	0%	83%	17%	0%
Top (F)	56%	27%	16%	36%	64%	0%	93%	7%	0%
Jacket (F)	52%	40%	9%	67%	33%	0%	78%	19%	3%
Skirt (F)	43%	33%	24%	43%	57%	0%	76%	24%	0%
Bag (F)	100%	0%	0%	100%	0%	0%	100%	0%	0%
Shoes (F)	6%	94%	0%	38%	12%	50%	18%	0%	82%
Suit (M)	54%	41%	5%	62%	38%	0%	57%	35%	8%
Jeans (M)	37%	54%	9%	54%	46%	0%	89%	11%	0%
Pants & Trousers (M)	34%	54%	11%	57%	37%	6%	69%	26%	6%
Shirt (M)	53%	35%	12%	41%	59%	0%	97%	3%	0%
T-shirt (M)	53%	29%	18%	38%	62%	0%	91%	9%	0%
Bag (M)	100%	0%	0%	100%	0%	0%	100%	0%	0%
Shoes (F)	9%	88%	3%	34%	3%	63%	25%	12%	63%

For similar product recommendations, the pseudo-list for each product was prepared as follows:

(a_1) randomly select a_1 products with the highest similarity score from the pool of candidates;

(a_2) randomly select a_2 products from the recommendation list originally suggested by the website; and

(a_3) randomly select a_3 products under the same product category and subcategory.

Since the number of candidate products under each product category was huge, as demonstrated in Table 1, we did not do exhaustive search using the proposed recommendation algorithm to identify a_1 products. Instead, we simply formed a candidate pool by sampling a subset of the available candidates, then calculated the similarity scores for the products within the candidate pool, and finally selected the candidates with the highest similarity scores within the pool. This candidate pool scheme is suggested because the fashion products are frequently updated for various reasons (e.g. stock replenishment); the scheme can strike a balance between computation expenses and recommendation effectiveness, making possible of online recommendation computations.

The pseudo lists, each with a maximum of 20 similar products (i.e. similar-lists), were prepared on the basis of the following considerations. The first half (50%) of similar recommendations were maximum ten products with the highest similarity score a_1; the second half of the recommendation list were 'noise', among which maximum five (25%) were originally recommended by ASOS a_2 and the other maximum five (25%) were products randomly selected from the same category and subcategory with the lowest similarity scores a_3.

The subjects were asked to select two to five similar products from a given similar-list that they consider having the highest similarity. Since the numbers a_1, a_2 and a_3 constituting the similar-list for a given product is not always a constant number of 20, we can then calculate the expected numbers being selected from each of the three sources. Based on each presented recommendation list, we can obtain expected number of products being selected by users from that

source. On the other hand, the actual number of products being selected from each source is known and recorded in wish-list. If more than expected numbers are actually selected from the source, it indicates that the products from that source are more similar to the given products. If less than expected numbers are actually selected by users, it implies that products from that source are of lower similarity.

Among the 523 valid similar results given by the 95 subjects (users), a total 10458 similar products were presented from which 1140 products were selected by the users. Table 2 shows the percentages of (I) manually selected number are as expected; (II) manually selected are more than expected from the source; and (III) manually selected number are less than expected. As shown in the table, the beyond expectation performance ratio (II) of our algorithm (source a_1) is obviously higher than that of random walk (source a_2), we can conclude that our algorithms are better than random walk. If what subjects choose on their owns are happened same as those suggested by the recommendation system, it is an indicator that the system can generate very effective recommendations that are comparable to humans. As shown in Table 2, 42% recommended list from our algorithms (source a_1) are manually selected by users, exceeding the expected performance (beyond expectation ratio II), which is similar to the ratio of human experts 46% (source a_2). It indicates that our algorithm are comparable to human experts, in some cases slightly under perform comparing to human experts, as revealed by (III) under expectation percentages. The possible reason is that we did not use exhaustive search to obtain the recommendation list in our algorithm, instead we make recommendation by a subset of candidates. In other words, what being suggested by human experts may be not within the candidate pool. It is therefore interesting to compare the average similarity scores of recommended products with the given items from sources a_1, a_2, and a_3. As shown in Table 3, similarity scores match the performance whether the products from the corresponding sources are being selected by users or not. As shown in Table 2 for cases that recommendations made by human experts outperform those recommended by our algorithms, the average similarity scores of human

experts are also higher than those being recommended by our algorithms in Table 3. It implies that similarity calculation given by Equation (2) reflects users preferences. We define the weighting parameter w_k in Equation (2) by fashion experts. Therefore, our similar recommendation method is demonstrated effective.

Table 3: Average similarity score.

	Source a_1: by algorithm	Source a_2: by website/experts	Source a_3: random walk
all	0.13	0.144	0.036
Cardigan (F)	0.194	0.213	0.033
Dress (F)	0.117	0.139	0.029
Top (F)	0.185	0.183	0.012
Jacket (F)	0.169	0.115	0.041
Skirt (F)	0.145	0.128	0.035
Bag (F)	0	0	0
Shoes (F)	0.012	0.1	0.036
Suit (M)	0.145	0.136	0.09
Jeans (M)	0.119	0.18	0.016
Pants & Trousers (M)	0.134	0.129	0.062
Shirt (M)	0.212	0.2	0.014
T-shirt (M)	0.185	0.223	0.037
Bag (M)	0	0	0
Shoes (F)	0	0.1	0.058

3.3. Mix-and-match recommendation results

For mix-and-match recommendations, the pseudo-list for each product was prepared as follows:

(b_1) randomly select b_1 number of products with the highest mix-and-match scores which were calculated by the algorithms, including two to three items from each matching category/subcategory;

(b_2) randomly select b_2 from the recommended list originally suggested by the websites, including as many as possible different categories/subcategories; and

(b_3) randomly select the noise b_3 from the matching category and subcategories.

The pseudo lists, each with a maximum of 20 mix-and-match products (i.e. matching-lists), were prepared as follows: The first half of the products were obtained by our mix-and-match algorithm (Section 2.5.2), which suggests matching items by integrating cross-media information learned from product text-based description and images. The second half was the original matching recommendations of ASOS, and some items randomly selected from the matching category and subcategory, which only text-based category information are used to make such recommendations.

In the experiment, subjects were asked to select from the matching-list items that can coordinate well with the selected item, covering as many categories as possible, usually ranging from three to four categories. For mix-and-match recommendations, the 95 subjects manually selected 1,259 from 6,674 mix-and-match recommendations.

In terms of mix-and-match results, matching items from a wide range of product categories are recommended to users, each recommended category has maximum three to four items to choose. Since, the number of suggested items in each category is likely less than five, we therefore calculated and compared the performance in Table 4. The comparison is shown in percentages being selected in comparison to the expected percentage: (I) as expected, (II) beyond expectation, and (III) less than expectation.

In general, 59% of the selected matching products are as expected, and 31% of the selected items are beyond the expected level of performance. The beyond expectation level is highest in most categories by our algorithms than other sources such as by expert and by random walk.

Table 4: Average fitness score for mix-and-match recommendations.

	Source b_1: by algorithm			Source b_2: by website/experts			Source b_3: random walk		
	(I)	(II)	(III)	(I)	(II)	(III)	(I)	(II)	(III)
all	59%	31%	10%	72%	28%	0%	88%	9%	3%
Cardigan (F)	33%	62%	6%	69%	31%	0%	73%	12%	15%
Dress (F)	45%	33%	22%	78%	22%	0%	56%	41%	3%
Top (F)	76%	11%	13%	55%	45%	0%	100%	0%	0%
Jacket (F)	50%	43%	7%	66%	34%	0%	72%	31%	7%
Skirt (F)	61%	35%	4%	76%	24%	0%	100%	0%	0%
Bag (F)	100%	0%	0%	100%	0%	0%	100%	0%	0%
Shoes (F)	100%	0%	0%	100%	0%	0%	100%	0%	0%
Suit (M)	19%	78%	3%	86%	14%	0%	81%	14%	5%
Jeans (M)	60%	20%	20%	63%	37%	0%	100%	0%	0%
Pants & Trousers (M)	71%	11%	17%	63%	37%	0%	100%	0%	0%
Shirt (M)	65%	24%	12%	59%	41%	0%	100%	0%	0%
T-shirt (M)	56%	35%	9%	56%	41%	0%	100%	0%	0%
Bag (M)	0%	100%	0%	100%	0%	0%	100%	0%	0%
Shoes (F)	100%	0%	0%	100%	0%	0%	100%	0%	0%

Similarly, we also calculate the average fitness score for mix-and-match recommendations from different sources in Table 5. Comparing Tables 4 and 5, mix-and-match fitness scores in general match with user preferences reflected in percentages of matching (I), beyond (II) and under (III) expectation. We will further research on mix-and-match recommendations, e.g. parameter and associating weight definitions of the fitness score, as similar to Equation (2); this will be explained in next section.

Table 5: Average fitness score for mix-and-match recommendations.

	Source b_1: by algorithm	Source b_2: by website/experts	Source b_3: noises
all	0.334473431	0.089249824	0.059112963
Cardigan (F)	0.458378303	0.129446652	0.146544222
Dress (F)	0.326804975	0.035471094	0.233599434
Top (F)	0.406452097	0.152695019	0
Jacket (F)	0.416889711	0.123326931	0.115664907
Skirt (F)	0.369651208	0.082951486	0
Bag (F)	0.282632966	0	0
Shoes (F)	0	0	0
Suit (M)	0.41011969	0.041407632	0.072093295
Jeans (M)	0.551834003	0.164798401	0
Pants & Trousers (M)	0.257579744	0.132751635	0
Shirt (M)	0.39058643	0.140912179	0
T-shirt (M)	0.408580519	0.162967515	0
Bag (M)	0.34217967	0	0
Shoes (F)	0	0	0

As shown, the algorithm suggested mix-and-match recommendations are more effective than those originally recommended by the website. Users prefer to choose what the algorithms suggested than recommended originally on ASOS website or those randomly suggested. However, it does show that some cate-

gories, e.g. the bags and shoes, the recommendation algorithm must be further improved in order to generate effective recommendations.

4. Conclusions and future work

In this paper, we have investigated the problem of fashion recommendations by integrating the textual mining and content-based information retrieval techniques. We have developed a new dataset by crawling fashion products from different online stores. We have formulated a method to make fashion recommendations with a number of developments. Firstly, we have developed a method that re-categorised products from different sources into the defined categories through text mining and semantic analysis. We have trained different deep CNN models to classify product image-types, parse input images into regions with semantic meaning, and recognise product fine-grained category. We have also proposed new methods to extract colour and texture features from segmented product regions. Lastly, we have proposed new algorithms to suggest similar and mix-and-match fashion items for any given products.

We have demonstrated by experiment that effective fashion recommendations can be obtained by integrating cross-media product information, which outperform than traditional text-based recommendations. A key contribution of this work is proposing similar and mix-and-match recommendation techniques based on human parsing features. Another contribution is developing a novel Pantone-based colour similarity descriptor. Finally, deep learning techniques are also adopted to extract the matching rules of fashion products.

The current mix-and-match recommendation method depends on if full-body images are available for analysis using human parsing techniques. If full-body images are not available, expert rules are used to obtain mix-and-match recommendations. It will be interesting to extend the current work of using human parsing technique to make mix-and-match suggestions by mining popular clothing coordination rules from a large number of full-body fashion images. The fashion product categories and attributes should be recognised and extracted

from fashion images. Statistical methods are used to analyse the mix-and-match rules and degree of matching. In addition, users profiles and behaviours (e.g. transaction data and reviews from users) should be used to obtain provide recommendations.

Acknowledgments

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 152161/17E). This work was also partially supported by The Innovation and Technology Fund (Grant No. ITS/253/15), The Hong Kong Polytechnic University (Grant No. G-YBRG and G-UA9L), Guangdong Provincial Department of Science and Technology (Project No. R2015A030401014) and Shenzhen Science and Technology Innovation Commission (Project No. JCYJ20170303160155330) in China. Wei Zhou was also supported by National Natural Science Foundation of China (Under grant No. 61602070).

References

[1] E. A., Us online retail sales will grow 57% by 2018, https://www.digitalcommerce360.com/2014/05/12/us-online-retail-sales-will-grow-57-2018/, [Online; accessed 15-Sept-2017] (2014).

[2] V. Jagadeesh, R. Piramuthu, A. Bhardwaj, W. Di, N. Sundaresan, Large scale visual recommendations from street fashion images, in: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2014, pp. 1925–1934.

[3] K. Yamaguchi, T. Okatani, K. Sudo, K. Murasaki, Y. Taniguchi, Mix and match: Joint model for clothing and attribute recognition., in: BMVC, 2015, pp. 51–1.

[4] X.-Y. Wang, B.-B. Zhang, H.-Y. Yang, Content-based image retrieval by integrating color and texture features, Multimedia tools and applications 68 (3) (2014) 545–569.

[5] S. Zhang, L. Yao, A. Sun, Deep learning based recommender system: A survey and new perspectives, arXiv preprint arXiv:1707.07435.

[6] K. Rahul, R. Agrawal, A. K. Pal, Color image quantization scheme using dbscan with k-means algorithm, in: Intelligent Computing, Networking, and Informatics, Springer, 2014, pp. 1037–1045.

[7] J. McAuley, C. Targett, Q. Shi, A. Van Den Hengel, Image-based recommendations on styles and substitutes, in: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2015, pp. 43–52.

[8] Y. Hu, X. Yi, L. S. Davis, Collaborative fashion recommendation: A functional tensor factorization approach, in: Proceedings of the 23rd ACM international conference on Multimedia, Brisbane, Australia, 2015, pp. 129–138.

[9] L. Liu, X. Du, L. Zhu, F. Shen, Z. Huang, Discrete binary hashing towards efficient fashion recommendation, in: International Conference on Database Systems for Advanced Applications, Gold Coast, Australia, 2018, pp. 116–132.

[10] F. Ricci, L. Rokach, B. Shapira, P. B. Kantor, Recommender systems handbook, Springer, 2015.

[11] S. Chandra, S. Tsogkas, I. Kokkinos, Accurate human-limb segmentation in rgb-d images for intelligent mobility assistance robots, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2015, pp. 44–50.

[12] D. ping Tian, et al., A review on image feature extraction and representation techniques, International Journal of Multimedia and Ubiquitous Engineering 8 (4) (2013) 385–396.

[13] G.-H. Liu, J.-Y. Yang, Content-based image retrieval using color difference histogram, Pattern Recognition 46 (1) (2013) 188–198.

[14] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 1, IEEE, 2005, pp. 886–893.

[15] G.-H. Liu, J.-Y. Yang, Z. Li, Content-based image retrieval using computational visual attention model, pattern recognition 48 (8) (2015) 2554–2566.

[16] O. Egozi, S. Markovitch, E. Gabrilovich, Concept-based information retrieval using explicit semantic analysis, ACM Transactions on Information Systems (TOIS) 29 (2) (2011) 8.

[17] G.-H. Liu, L. Zhang, Y.-K. Hou, Z.-Y. Li, J.-Y. Yang, Image retrieval based on multi-texton histogram, Pattern Recognition 43 (7) (2010) 2380–2389.

[18] X. Wang, T. Zhang, D. R. Tretter, Q. Lin, Personal clothing retrieval on photo collections by color and attributes, IEEE Transactions on Multimedia 15 (8) (2013) 2035–2045.

[19] X. Song, F. Feng, J. Liu, Z. Li, L. Nie, J. Ma, Neurostylist: Neural compatibility modeling for clothing matching, in: Proceedings of the 2017 ACM on Multimedia Conference, ACM, 2017, pp. 753–761.

[20] X. Song, F. Feng, X. Han, X. Yang, W. Liu, L. Nie, Neural compatibility modeling with attentive knowledge distillation, in: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '18, ACM, New York, NY, USA, 2018, pp. 5–14.

[21] L. Nie, X. Wang, J. Zhang, X. He, H. Zhang, R. Hong, Q. Tian, Enhancing micro-video understanding by harnessing external sounds, in: Proceedings of the 2017 ACM on Multimedia Conference, ACM, 2017, pp. 1192–1200.

[22] L. Nie, L. Zhang, L. Meng, X. Song, X. Chang, X. Li, Modeling disease progression via multisource multitask learners: A case study with alzheimers disease, IEEE Trans. Neural Netw. Learning Syst 28 (7) (2017) 1508–1519.

[23] L. Nie, L. Zhang, Y. Yang, M. Wang, R. Hong, T.-S. Chua, Beyond doctors: Future health prediction from multimedia and multimodal observations, in: Proceedings of the 23rd ACM international conference on Multimedia, ACM, 2015, pp. 591–600.

[24] L. Nie, L. Zhang, M. Wang, R. Hong, A. Farseev, T.-S. Chua, Learning user attributes via mobile social multimedia analytics, ACM Transactions on Intelligent Systems and Technology (TIST) 8 (3) (2017) 36.

[25] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.

[26] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1520–1528.

[27] P. Jing, Y. Su, L. Nie, X. Bai, J. Liu, M. Wang, Low-rank multi-view embedding learning for micro-video popularity prediction, IEEE Transactions on Knowledge and Data Engineering 30 (8) (2018) 1519–1532.

[28] W. Di, C. Wah, A. Bhardwaj, R. Piramuthu, N. Sundaresan, Style finder: Fine-grained clothing style detection and retrieval, in: Proceedings of the IEEE Conference on computer vision and pattern recognition workshops, 2013, pp. 8–13.

[29] P. Jing, Y. Su, L. Nie, H. Gu, J. Liu, M. Wang, A framework of joint low-rank and sparse regression for image memorability prediction, IEEE Transactions on Circuits and Systems for Video Technology.

[30] K. Yamaguchi, M. Hadi Kiapour, T. L. Berg, Paper doll parsing: Retrieving similar styles to parse clothing items, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 3519–3526.

[31] Y. Kalantidis, L. Kennedy, L.-J. Li, Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos, in: Proceedings of the 3rd ACM conference on International conference on multimedia retrieval, ACM, 2013, pp. 105–112.

[32] L. Zheng, S. Wang, Z. Liu, Q. Tian, Packing and padding: Coupled multi-index for accurate image retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1939–1946.

[33] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556.

[34] Y. Zhou, P. Mok, A part-detection based and crfs embedded deep neural network for human parsing, in: 3rd International Conference and Expo on Computer Graphics and Animation, Las Vegas, USA, 2016.

[35] X. Liang, S. Liu, X. Shen, J. Yang, L. Liu, J. Dong, L. Lin, S. Yan, Deep human parsing with active template regression, IEEE transactions on pattern analysis and machine intelligence 37 (12) (2015) 2402–2414.

[36] N. Carlstadt, Pantone fashion home interiors : Cotton planner : New colors, Pantone LLC, 2015.

[37] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.