Sensing and Imaging (2019) 20:30
https://doi.org/10.1007/s11220-019-0249-8

ORIGINAL PAPER

Structuring Description for Product Image Data with Multilabel

Yong Dai[1,3] · Yi Li[2] · Li-Jun Liu[2]

Received: 2 April 2019 / Revised: 26 April 2019 / Published online: 5 July 2019

Abstract

Existing product data, presented as description documents, product images, etc., are very significant references for designing a new proposal. However, traditionally, the data collected manually only contain the product images without description documents. Obtaining reasonable description of product is a challenging task due to the high cost and labor-intensive process of human annotation. In this paper, a new approach is introduced to solve this problem and improve the efficiency of description by exploiting the potential information of product images. We propose a robust framework with multi-label learning to annotate each product image with several labels, which makes a brief description of the product from different aspects. An efficient method is proposed to accomplish product data acquisition, arrangement and analysis task to construct new data in images collection step. Furthermore, for the newly structured data, robust algorithms are employed to accomplish the data processing. Based on the processed data, automatic and semi-automatic multi-label annotation methods are applied to generate the labels for product images. We present the prediction results by several state-of-the-art multi-label learning classifiers based on the features extracted from different convolution neural framework. The results are evaluated by effective measures to validate the quality of the labeling result, as well as the predictive performance with respect to the number of the training examples. Experimental results indicate that the quality of the annotation is reasonable, and the proposed method can achieve excellent prediction accuracy based on the annotated data, generating satisfactory description for the products.

Keywords Design knowledge · Description documents · Product image data · Data preprocess · Multi-label learning

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11220-019-0249-8) contains supplementary material, which is available to authorized users.

✉ Yi Li
2012171@hnu.edu.cn

Extended author information available on the last page of the article

1 Introduction

Product design knowledge refers to the knowledge generated in the process of product design, which is coming from existing product data. Product data are very significant references for designing a new proposal and usually presented as description documents, product images, etc. After reviewing and analysing the previous product example, designers could find the features and styles from description documents and product images created by skilled designers to seek the inspiration and then propose the initial design scheme. So both description documents and product images should be acquired and analysed before designing a new proposal.

It is reported that the image projected by a product depends to a large extent upon its physical form, and certain critical features of the product images could affect a consumer's psychological response to the product [23]. So existing product images are very crucial carrier for product design knowledge and can also be the references for designing a new proposal.

On traditional product design task, they are always collected at the beginning of the product design task. However, product image data were usually collected manually in traditional design work which costs about even 27% of the whole design task according to some reports. Most importantly, the data collected by designers cannot cover most of the product images. With rapidly update in the product design environment, it is difficult to grasp the current product design trends and design concept with only a small number of product data. Besides, designing with only a small number of product data is easily resulting in the appearance of the same design and bad law and economic consequences. The rapid growth of product design has called for the intelligent methods to accomplish image collecting task. At present, the enterprise official websites and shopping platforms offer a huge amount of product data. Instead of a manual process, advanced big data and deep learning technology could be used for mass product data acquisition and analysis [31, 44].

Another crucial carrier for product design knowledge is description documents. However, most of collected data just contain the product images without description documents or sometimes description texts can be obtained but are chaotic, complex and useless for product design knowledge. Description documents are usually used when demonstrating new products to the market. So they mainly contain some attribute data of the product, design style and design elements which could semantically describe the products. In this case, a new idea is proposed to structure the semantic descriptio with multi-label learning since multi-label technology could endow several characteristics such as time, style and function. We claim that multi-label learning can provide valuable information on product data, especially in the case of description documents in which each product image can have its corresponding semantic description.

Departing from traditional single-label classification, in multi-label learning, each sample is associated with multiple labels simultaneously. However, in product design field, most of the time, product images only contain the products themselves, that means the images only have one label in traditional annotation norm. we argue that there are two differences between traditional multi-label case and the case in this paper.

First, for multi-label annotation, labels represent certain object presented in images in traditional case. So an annotator is asked to determine whether the images have or not for each label. In our case, attribute data of the product and the elements of industrial design are also used to better describe the products. So it is hard to determine whether the product images have or not for these labels. On the contrary, annotator need to determine the images should be labelled with this or that for each label. For example, '1' means one image has the current label,'0' means not. Just as the Sect. 4 present, binary '0' and '1' represent the color numbers less or equal than and greater than 2 for the label of the value of HSV color components, respectively.

Second, for multi-label prediction, it has a set of labels where multiple labels may be assigned to each object where each object only have one label in the traditional case. So in this way all the labels corresponding to the object can exactly belong to one subset ('object'). And the features corresponding to each object are fairly different. In our case, we have 20 labels from five subsets of labels (category, chromatic or achromatic color, color numbers, value of HSV color components and morphology) where exactly labels from five subsets are assigned to only one object. Most importantly, only one sutset of labels corresponds to the category of object, so constructing the framework to model the relationship between the labels and the only one object would be a harder problem than the problem in traditional multi-label case.

Motivated by this, we present a pre-processing procedure to accomplish product data acquisition, sorting and analysis to construct new data with the goal of shortening the period of data collection. And based on our newly structured data, we propose the application of multi-label learning to supplement the description documents for product design knowledge.

To the best of our knowledge, this is the first work which applies multi-label research in the product design field. The key contributions of our work are as follows:

- The construction of the data for the product design field, which can avoid the high cost and labor-intensive process and greatly shorten the design period and improve efficiency in images collection task.
- An efficient procedure for image processing to improve the quality of the large scale data collected from the Internet, and this procedure can be easily generalized to add more categories to augment or update the data.
- Multi-label learning to convert a huge pile of images into structured data, which can also be used for multi-label semantic learning.

The rest of the paper is structured as follows. Section 2 gives the related work on product image data and multi-label learning methods. Section 3 presents how we obtain the images and how we pre-process them, including the removal of "cluttered " and duplicate images. Section 4 is devoted to the introduction of multi-label annotation based on this data. Conclusions and future work are finally reported in Sect. 5.

2 Related Work

With the development of computer vision, increasing number of researchers pay attention to the understanding of visual scenes. Methods were therefore proposed to recognize the objects in the scene images and characterize the relationship between objects and corresponding scenes. To copy with the requirement for advancing the object-related research, several object classification and detection data [8, 9, 28, 39] were constructed and played an important role in the recent breakthroughs in both object classification and detection research. However, as the research moves along, it is found that there may exists more than one object in scene images. So traditional single label classification cannot meets requirements for fully describing the scene images. Thus, multi-label classification research became one of the hotest topics in recent years. Similarly, several multi-label annotation data such as scene [2], NUS-WIDE [5], Microsoft COCO [20], etc. were also presented as new benchmarks to test the performance of the different multi-learning methods.

Most data contain a large number of scenes and the objects that commonly occur in them or some nature images. But no data were constructed in the product design field. Product data are very important references for designing a new proposal and are usually collected manually in traditional design work which usually costs much time, labor and money. Recently, designers can download images from some websites which provide the platforms for designers to share product images of original design or existing products. But there are problems such as mixed background and low relevance of image retrieval.

The multi-label learning framework has attracted considerable attention in the literature over the last decade due to its numerous real-world applications [29, 32]. Traditionally, it is applied in text [17, 19], audio [35, 41], image classification [16], and even image segmentation [25]. Departing from traditional single-label classification, in multi-label learning, each sample is associated with multiple labels simultaneously. One of the primary goals of multi-label learning is to understand the objects categories or scenes and objects that commonly occur in them. Objects understanding involves multiple tasks, one of which is recognizing what objects are present in one scene image. During the past decade, significant amount of progresses have been made towards this emerging machine learning paradigm [43].

Traditionally, for multi-label annotation, labels represent certain objects present in images. So an annotator is asked to determine what objects are present for each label. But in product design field, most of the time, product image only contain one object, i.e., the product itself, which means the images only have one label in traditional annotation norm. To the best of our knowledge, there is no work that considers about this situation. Inspired by the missing of description documents when collecting product data, we propose the multi-label learning research in the product design field.

3 Data Acquisition and Processing

In this section, the proposed method for image acquisition and pre-processing will be presented in detail. To get as many samples as possible for the products of each category, images are crawled from several big online shopping websites, which are bigger than enterprise official platforms. One key contribution of this work lies in fusing multi-source data and preprocessing the data with robust methods to ensure the integrality of the data.

3.1 Multi-source Data Fusion

Web crawler is a program that can automatically grab information from webs based on certain rules. So researchers usually utilize this technology to obtain specific data existing in certain webs. To ensure the integrality of the data, we crawl images from several big online shopping websites. However, different shopping websites may have the same products. It is inevitable that there exist duplicate products in the multi-source data. Fortunately, shopping websites provide identification number for each product in each corresponding specifications. Hence, when product images are crawled, the corresponding identification number is also crawled simultaneously.

More than 160,000 images are crawled from several big online shopping websites. After crawler task is finished, identification numbers for products are first read and compared between each other, the product images which are with the same identification numbers will be put into the same folder. Then all folders are gathered according to the category.

3.2 Data Arrangement

It is observed that: (1) some images have different size, (2) some images have cluttered background which are useless to be the reference for design work, (3) some images are same. So firstly, the images which have the size of 200 × 200 pixel are remained, others whose sizes are inappropriate or have bad length-width ratios are resized to the size of 200 × 200 pixel.

Next step is to remove the cluttered images. At the beginning of this step, the frequency-tuned salient region detection [13] is used to get salient objects since it could achieve the best results among most of saliency detection algorithms [1, 4, 45]. Then edge detection algorithm is used to get the edge map based on salient objects images, after which, feature extraction is applied to the inter-edge area and outer-edge area of the salient objects. Feature judgement is used to classify whether the backgrounds of images are pure and white or not, the images with poor backgrounds or unclear edges of products are “cluttered ” and are selected to be deleted, on the contrary, the images where products sit on the central zone of images and the background is pure and white are remained. Taking smart watch as an example, Fig. 1 shows the examples of “cluttered ” (a) and “pure” (b) images.

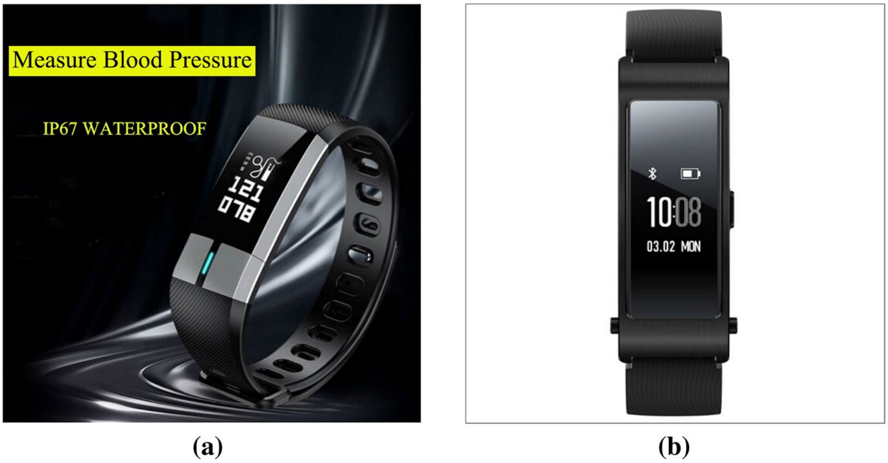

Fig. 1 An example of "cluttered " and "pure" images. **a** the "cluttered " image with poor background or unclear edge of product; **b** "pure" image with pure white background and clear edge of product

To get rid of data redundancy, the comparison of similarity based on dhash [42] is put forward to remove the duplicated images. Firstly, perceptual hash algorithm is used to computer dhash value for the remained images after "cluttered " removal. Then an image is randomly selected as sample A and the other images are selected in order as sample B. Hamming distance (*L*) between sample A and B with the threshold *T* are compared to determine whether to remove sample B or remain it. The Hamming distance (*L*) can be regarded as the dissimilarity between two samples. The threshold *T* is set as 0 since we want to remove the duplicates of the same images. This flow is circled until sample B is the last one in dhash list. Next, sample A is remained in the final data, and reselecting another one image as sample A after dhash list updated. Finally, the full duplicate removal flow is circled until sample A traverse all the samples. The formal steps are achieved automatically using data analysis algorithms coded by Python, decreasing several hundred worker hours. This procedure for image pre-process can be easily employed to add more categories to augment the data. And the flow diagram of the image pre-processing is presented in Fig. 2.

The remain set contains 64,893 images including a total of 16 categories of daily electronic products with more than 4 thousand brands, 76.3% of the products have more than 5 images including front view, top view, left view and some local view in detail. Hence, new folders are set up to classify the products images according to the views, and put the images with the same view of products into the same folders. Figure 3 gives the categories of samples and their corresponding amounts and brand numbers. The data will be augmented in order to make data become larger and contain more categories, the complete data set will be available online as soon as possible.

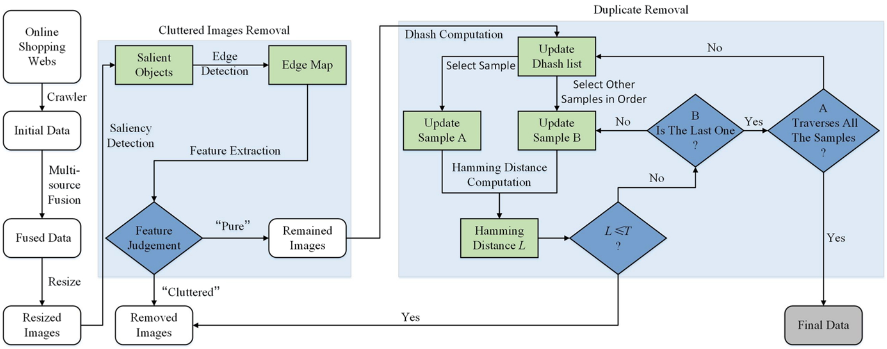

Fig. 2 The flow of the image pre-processing

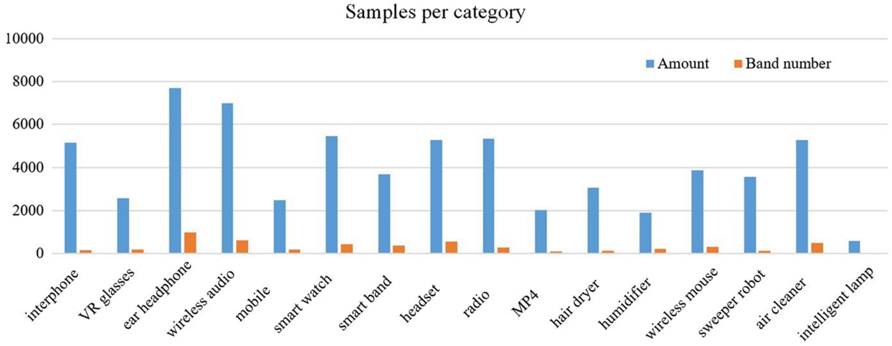

Fig. 3 The categories of samples and their corresponding amounts and brand numbers

4 Multi-label Semantic Learning

4.1 Label Analysis

Annotating almost 65 thousand images is an extremely time consuming work and is nearly impossible to manually annotate all images. In the meantime, due to subjective feelings, they may not grasp the meaning of some images and fail to make critical annotation. So several multi-label annotation algorithms are strategically applied to assist annotators to complete this task to minimize the cost and increase the precision. However, this semi-manual annotation of ground-truth still has some shortcomings to a certain degree, since annotator inescapably has their own subjective feeling that may lead to the unsatisfactory training data for multi-label learning, in this case, the precise and recall of algorithms will be lower than the expectation. To address these shortcomings, convincing multi-label algorithms could be adopted to reduce the effect of subjective feelings as far as possible. What's more, label selection is a crucial way to verify the effectiveness of algorithms, and subjective labels are the best choices.

Below, to verify the effectiveness of multi-label learning algorithms and to facilitate experimentation, we firstly select a set of important and basic labels which are picked from attribute data of the product and the elements of industrial design. With the addition of the categories of products, the number of labels is extended to 20.

1. Category

   As mentioned before, the data contain a total of 16 categories of daily electronic products with more than three thousand brands. Although the brand is one part of the attribute data of the product, there are too many brands included in the data. We just put the categories rather than brands into the annotation task.
2. Chromatic or achromatic color

   Achromatic color is a combination of white, black, and a variety of degree of gray that are formed from white and black. According to certain change rule, achromatic color can be arranged in a series, from white gradient to light gray, medium gray, dark gray and black, which is called as black and white series. On the contrary, chromatic is a combination of red, orange, yellow, green, blue and purple.
3. Color numbers

   When designing products, designers usually need to consider the color assortment. So the number of colors is also a design element that should be taken into consideration. Since there are not only two kinds of color numbers due to the wide variety of products, a threshold is needed to split it to two parts to fit the binary code for each label. The threshold is set to ‘2’ on the basis of statistics result, so the binary ‘0’ and ‘1’ represents the color numbers less or equal than and greater than ‘2’, respectively.
4. Value of HSV color components

   Value indicates the bright degree of color. For object color, this value is tied to the transmittance or reflectance of an object. The range is usually 0% (black) to 100% (white), similar to the color number, after setting a threshold, so the binary ‘0’ and ‘1’ represents the value less or equal than and greater than the threshold, respectively.
5. Morphology

   Morphology is another kind of design element that is usually considered when designing a product. Common morphology usually contain two kinds such as circular type and rectangle type.

4.2 Image Annotation

Due to the characteristic of objective labels, these labels should be able to be annotated automatically. But sometimes, images of some categories may effect the results. Taking mobile as an example, the picture shown in mobiles screens may decrease the precision of annotation for the label of color numbers and Chromatic or achromatic color. Hence, the annotation task is divided into two parts, one part is annotated by code, and the other one part is annotated by annotator with the help of human computer interaction (HCI) application.

1. Automatic annotation

   For the labels of category and value of HSV color components, programs are written to annotate the labels automatically. For the labels of category, it's easy to accomplish the task since the data are managed into their corresponding folders of category when crawled from webs. For the labels of value of HSV color components, we referred to the pre-processing step which is discussed in Sect. 2. Saliency detection is applied first to get the region of objects, and then the value of the certain region of the image could be obtained. To keep the balance of label '0' and '1', statistics suggest that the numbers of samples with two kinds of labels are approximate when the threshold is set to 80%.
2. Semi-manual annotation

   For the labels of chromatic or achromatic color, color numbers and morphology, HCI application is designed to assist the annotation task, an annotator is asked to determine to annotate '0' or '1' for the images for each label.

   Next, we briefly introduce the semi-manual annotation procedures for annotating the multiple labels. By the way, in this paper, annotation is different to the traditional annotation task for the last four labels. In the traditional multi-label annotation, an annotator is asked to determine whether the images have or not for each label. But in this paper, annotator has to determine the images should be labelled with this or that for each label. For instance, '1' means one image has the current label,'0' means not. Just as the formal present, the binary '0' and '1' represents the color numbers less or equal than and greater than '2', respectively.

As shown in Fig. 4, the annotators manually view and annotate about a quarter of all images using a hierarchical labeling approach [7, 15] for each category [5, 20], these portion of annotated images are used as the training data to perform multi-label induction on the remaining unlabeled images, then semi-manual annotation is applied. Two state-of-art multi-label algorithms are used to generate the annotation for each unlabeled image by using VGG_16 [30] feature to train data, these two algorithms could achieved the best result compared to other algorithms [18]. RAkEL–DT represents the ensemble approach of RAndom k-labELsets [36] and Decision Trees [6], ECC–DT is the ensemble approach of classifier chains [26] and Decision Trees [6]. Then the corresponding annotation results *L(R-D)* and *L(E-D)* from RAkEL–DT and ECC–DT algorithm are compared in the judgement procedure.

If these two annotation results are equal for the same inputs, the equal annotation will be output to the training data to extend the training data, otherwise, the images with different annotations will be annotated manually again. If the annotator sees a certain concept exist in the image, label it as positive; if the concept does not exist in the image, or if the annotator is uncertain on whether the image contains the concept, then label it as negative [5].

4.3 Experiments Setting and Evaluation

As for state-of-the-art multi-label algorithms, a mass of researches by the machine learning community have provided a large number of multi-label

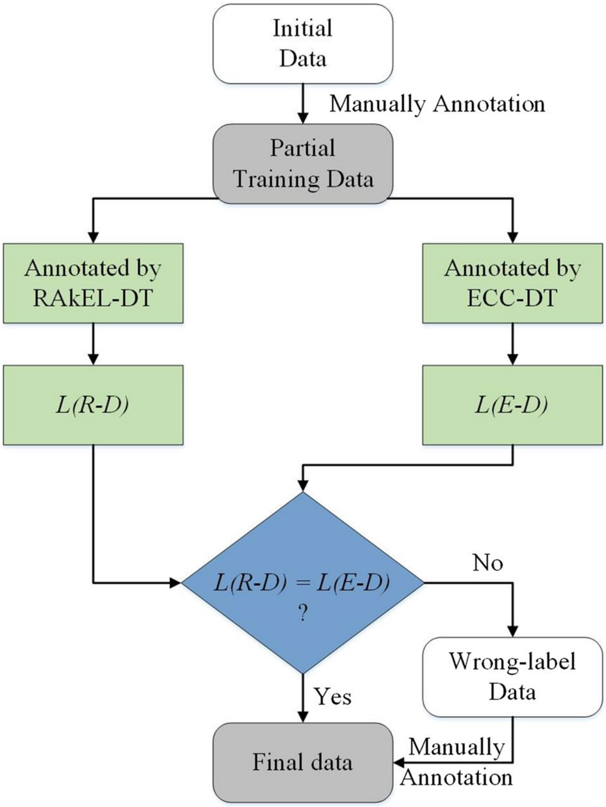

Fig. 4 The flow of multi-label annotation

approaches [11, 43]. These approaches can be broadly divided into three categories according to different theories, including problem transformation, algorithm adaptation and ensemble approaches [22]. The existing data [5, 20] used for multi-label learning usually applied kNN based algorithms to evaluate the data since the kNN based algorithms are better than most of the problem transformation methods, however, kNN based algorithms which is included in algorithm adaptation methods sometimes could not get the best performance compared to the ensemble approaches [18]. Due to these properties, ensemble approaches are very attractive.

We finally decided to choose two ensemble approaches including EA(LP–RF) and EA(BR–SVM). EA(LP–RF) represent the ensemble approach of Random Forests [3] under a Label Powerset [37] multi-label classifiers, and EA(BR–SVM) means the ensemble approach of Support Vector Machine (SVM) [38] under a Binary Relevance (BR) [21] multi-label classifiers. To conduct the experiments, we consider the algorithms implementations provided by Wroclaw University of Technology [34]. Random selection of subsets is likely to negatively affect the ensemble's performance since the chosen subsets may not cover all labels or inter-label relationship [27]. Note that both approaches partition each label subspace using fast greedy community detection methods on a label co-occurrence graph [10] instead of overlapping RAndom k-labELsets [36]. Further details can be seen in [34].

However, these two ensemble approaches need to learn from the extracted features instead of images. CNN based algorithms can effectively extract features from original images, so we extract inception_V1 [33], inception_V2 [14], ResNet_V1-50, ResNet_V1-101 [12], VGG_16 and VGG_19 [30] features to evaluate the performance of the annotation. We randomly select 2 thousand samples from each category except intelligent lamp to avoid sample disproportion, and make a split of the training to testing samples in the order of 8:2, the overview of the performance is given in Tables 1 and 2.

Assuming $T = \left\{ \left(x_i, Y_i \right) | i = 1 \ldots n \right\}$ be the multi-label data with n examples, where i means the ith sample. Note that in the following formulas, for a test sample x, Z_x is defined as its predicted set of labels, and $r_x(\lambda)$ means the associated ordered rank for the label λ. Based on this, the measures we choose to evaluate the performance are as follows:

Table 1 Performance (mean ± SD) of each ensemble approaches over ten measures

Multi-label learning algorithms	Measures	Features extraction models		
		Inception_V1	Inception_V2	Resnet_V1_50
EA(LP + RF)	Coverage↓	4.558 ± 0.005	3.856 ± 0.031	3.866 ± 0.040
	Average precision↑	0.847 ± 0.001	0.895 ± 0.001	0.902 ± 0.002
	Ranking loss↓	0.154 ± 0.001	0.107 ± 0.001	0.098 ± 0.003
	Hamming loss↓	0.060 ± 0.001	0.041 ± 0.001	0.037 ± 0.001
	Micro precision↑	0.874 ± 0.001	0.917 ± 0.002	0.924 ± 0.001
	Micro recall↑	0.863 ± 0.001	0.901 ± 0.002	0.912 ± 0.003
	Micro F_score↑	0.868 ± 0.001	0.908 ± 0.001	0.918 ± 0.002
	Macro precision↑	0.851 ± 0.001	0.901 ± 0.001	0.913 ± 0.002
	Macro recall↑	0.812 ± 0.002	0.876 ± 0.001	0.887 ± 0.004
	Macro F_score↑	0.831 ± 0.001	0.888 ± 0.001	0.900 ± 0.003
EA(BR + SVM)	Coverage↓	5.779 ± 0.028	4.632 ± 0.001	6.043 ± 0.011
	Average precision↑	0.829 ± 0.003	0.876 ± 0.001	0.803 ± 0.001
	Ranking loss↓	0.201 ± 0.003	0.143 ± 0.002	0.213 ± 0.001
	Hamming loss↓	0.054 ± 0.001	0.038 ± 0.001	0.064 ± 0.001
	Micro precision↑	0.952 ± 0.001	**0.960 ± 0.001**	0.911 ± 0.001
	Micro recall↑	0.806 ± 0.004	0.867 ± 0.002	0.801 ± 0.001
	Micro F_score↑	0.873 ± 0.002	0.911 ± 0.001	0.852 ± 0.001
	Macro precision↑	0.966 ± 0.001	**0.974 ± 0.001**	0.826 ± 0.001
	Macro recall↑	0.713 ± 0.004	0.803 ± 0.004	0.723 ± 0.001
	Macro F_score↑	0.820 ± 0.002	**0.980 ± 0.022**	0.771 ± 0.001

For each metric, ↑ indicates "the higher, the better", whereas ↓ indicates "the lower, the better" and the bold indicates the best value

Table 2 Performance (Mean ± SD) of each ensemble approaches over ten measures

Multi-label learning algorithms	Measures	Features extraction models		
		Resnet_V1_101	VGG_16	VGG_19
EA(LP + RF)	Coverage↓	3.907 ± 0.024	3.672 ± 0.005	**3.659 ± 0.008**
	Average precision↑	0.900 ± 0.001	**0.917 ± 0.001**	0.916 ± 0.001
	Ranking loss↓	0.098 ± 0.001	**0.079 ± 0.001**	0.082 ± 0.001
	Hamming loss↓	0.038 ± 0.001	**0.031 ± 0.001**	0.032 ± 0.001
	Micro precision↑	0.921 ± 0.001	**0.935 ± 0.001**	0.934 ± 0.001
	Micro recall↑	0.911 ± 0.001	**0.929 ± 0.001**	0.923 ± 0.001
	Micro F_score↑	0.916 ± 0.001	**0.932 ± 0.001**	0.929 ± 0.001
	Macro precision↑	0.916 ± 0.001	0.929 ± 0.001	**0.930 ± 0.001**
	Macro recall↑	0.887 ± 0.002	**0.911 ± 0.001**	0.906 ± 0.001
	Macro F_score↑	0.901 ± 0.001	**0.920 ± 0.001**	0.918 ± 0.001
EA(BR + SVM)	Coverage↓	5.865 ± 0.036	**4.327 ± 0.005**	4.372 ± 0.011
	Average precision↑	0.812 ± 0.002	**0.895 ± 0.001**	0.894 ± 0.001
	Ranking loss↓	0.198 ± 0.002	**0.108 ± 0.001**	0.113 ± 0.001
	Hamming loss↓	0.061 ± 0.001	**0.033 ± 0.001**	0.034 ± 0.001
	Micro precision↑	0.912 ± 0.001	0.954 ± 0.007	0.954 ± 0.001
	Micro recall↑	0.813 ± 0.002	**0.901 ± 0.001**	0.895 ± 0.001
	Micro F_score↑	0.859 ± 0.001	**0.927 ± 0.001**	0.924 ± 0.001
	Macro precision↑	0.830 ± 0.001	0.962 ± 0.001	0.956 ± 0.001
	Macro recall↑	0.745 ± 0.003	**0.846 ± 0.001**	0.840 ± 0.001
	Macro F_score↑	0.785 ± 0.002	0.900 ± 0.001	0.894 ± 0.001

For each metric,↑ indicates "the higher, the better", whereas ↓ indicates "the lower, the better" and the bold indicates the best value

(a) Coverage (C) calculates the average distance for all of the relevant labels of the example for the ranked label list.

$$C = \frac{1}{n}\sum_{i=1}^{n}\max_{\lambda\in Y_i} r_{x_i}(\lambda) \text{ - } 1 \tag{1}$$

(b) Average precision (AP) shows the percentage of labels ranked above a special relevant label.

$$AP = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\left|Y_i\right|}\sum_{\lambda\in Y_i}\frac{\left|\left\{\lambda' \in Y_i : r_{x_i}\left(\lambda'\right) < r_{x_i}(\lambda)\right\}\right|}{r_{x_i}(\lambda)} \tag{2}$$

(c) Ranking loss (Rl) evaluates the average part of label pairs that are false ordered.

$$Rl = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\left|Y_i\right|\left|\bar{Y}_i\right|}\left|G_i\right| \tag{3}$$

where G_i represents $\left\{\lambda', \lambda'' \in Y_i : r_{x_i}\left(\lambda'\right) < r_{x_i}\left(\lambda''\right)\right\}$.

(d) Hamming loss (Hl) reports the percentage of example-label pairs in which label is predicted incorrectly.

$$Hl = \frac{1}{n}\sum_{i,j=1}^{n} 1\left(Y_i \neq Y_j\right) \tag{4}$$

where $1(x)$ is indicator function.

(e) Micro and macro measures.

$$\text{Micro-precision} = \frac{\sum_{j=i}^{m} TP_{\lambda_j}}{\sum_{j=i}^{m} TP_{\lambda_j} + \sum_{j=i}^{m} FP_{\lambda_j}} \tag{5}$$

$$\text{Micro-recall} = \frac{\sum_{j=i}^{m} TP_{\lambda_j}}{\sum_{j=i}^{m} TP_{\lambda_j} + \sum_{j=i}^{m} FN_{\lambda_j}} \tag{6}$$

$$\text{Macro-precision} = \frac{1}{m}\sum_{j=1}^{m} \frac{TP_{\lambda_j}}{TP_{\lambda_j} + FP_{\lambda_j}} \tag{7}$$

$$\text{Macro-recall} = \frac{1}{m}\sum_{j=1}^{m} \frac{TP_{\lambda_j}}{TP_{\lambda_j} + FN_{\lambda_j}} \tag{8}$$

where *TP*, *TN*, *FP*, *FN* means true positives, true negatives, false positives and false negatives, respectively, m is the number of labels. F_measure can be calculated as the harmonic mean between precision and recall, see [40] for further details.

For each feature, five different cross validation experiments are conducted for each algorithm, and the average value is calculated as the final result. Considering the ensemble approaches of Random Forests under a Label Powerset multi-label classifiers, VGG_16 feature achieves the best performance for all the measures except coverage and macro precision, for the measure of average precision, the scores for most of the features are approximately or above 90%. For the ensemble approaches of Support Vector Machine (SVM) under a Binary Relevance (BR) multi-label classifiers, VGG_16 feature still has a clear advantage for most of the measures except Micro precision, Macro precision and Macro F_score, The best score for these three metrics are achieved by the inception_V2 feature, and the scores are even high than 96%, suggesting that false positives are very small and true positive are fairly high. Analyzing the entire form, we observe that almost all the features based on these two ensemble approaches could achieve quite high score, so we can argue that the quality of the annotation is reasonable and satisfactory. Next step, we examine the

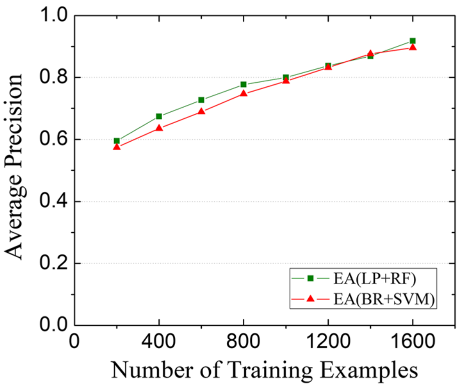

Fig. 5 Average precision with regard to the number of training samples for two ensemble algorithms

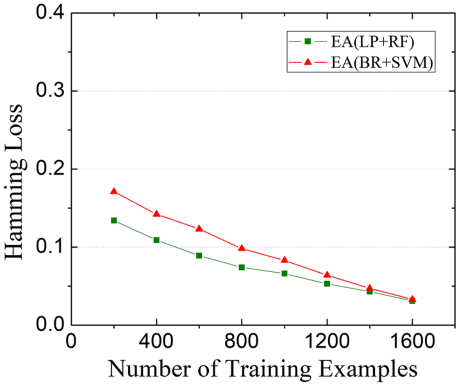

Fig. 6 Hamming loss with regard to the number of training samples for two ensemble algorithms

predictive performance with respect to the number of training samples. This is a critical parameter since it is directly related to the cost and manpower required for the classification and understanding of newly acquired product images. We consider a varying number of training samples ranging from 200 to 1600 for each category, averaged over 8 realizations, and the testing samples still remain at 400 for all realizations.

Different to the former experiment, this step mainly focuses on the number of training samples, so only one feature is enough. Tables 1 and 2 suggest that VGG_16 feature has a clear advantage compared with other features, so it is selected to be the basic feature for training samples. Additional parameters are set same with the former except the number of training samples. Figures 5, 6, 7 and 8 gives complex interactions between training data size and classification performance with regard to the number of training samples.

Figure 5 gives the interaction between average precision and the number of training samples for these two ensemble algorithms. It is observed that, the average precision achieves almost 0.6 when the number of testing samples is twice as large as training samples from the beginning. Then the average precision for both algorithms increases smoothly and steadily before the number of training samples

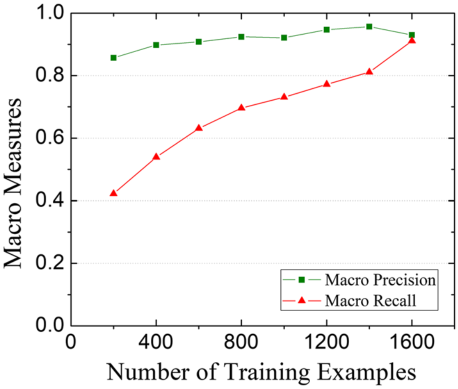

Fig. 7 Macro precision and recall for EA(LP + RF) with regard to the number of training samples

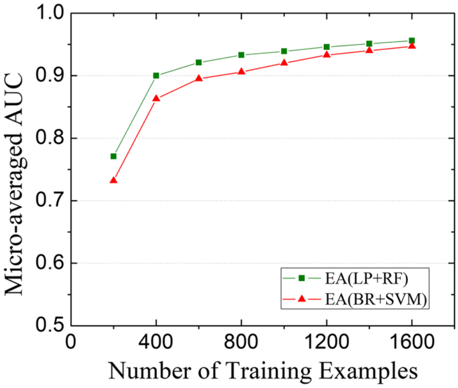

Fig. 8 Micro-averaged AUC with regard to the number of training samples for two ensemble algorithms

reaches to 1400, after which it increases slower for EA(BR + SVM) algorithm than EA(LP + RF) algorithm. The BR approach is the lack of consideration for label correlations while LP methods can capture interrelationships among labels, so it may lead to under- or overestimation of the active labels or the identification of multiple labels that never co-occur for EA(BR + SVM) algorithm when training set size is bigger than 1400.

Figure 6 shows the variation results of Hamming loss. This metric is highly representative since the Hamming loss belongs to the example-based metrics and can give us an overall intuition of the misclassified object-label pairs. Opposite to average precision, Hamming loss for both algorithms drops with the number variation of training samples. It decreases steadily for EA(BR + SVM) algorithm and EA(LP + RF) algorithm due to more and more sufficient training and reaches almost equal when the number of training samples reaches to 1600.

Tables 1 and 2 illustrate that the macro and micro precision and recall for EA(BR + SVM) algorithm vary more widely than EA(LP + RF) algorithm to some extent. To figure out the interactions between training data size and precision and recall, further experiment is conducted. Figure 7 just gives the Macro precision and recall for the EA(LP + RF) algorithm since macro measures focus more on classes instead of samples compared with micro measures, regarding less of how many documents belong to it [40]. We observe that the macro precision always achieves at a high score and increases slowly from 200 to 1600 for the number of training samples. While the macro recall changes greatly from approximately 0.4 to 0.9, increasing sharply with the number of training samples growth. This illustrates that the number of FP is fairly smaller than FN compared to TP, which may due to insufficient training for SVM when training set size is small.

More specifically, we consider the area under the curve (AUC) metric which is calculated from the receiver operating characteristic (ROC) curve. The AUC score describes the overall quality of performance, independently of individual threshold configurations regarding specific trade-offs between TP and FP [24]. Different to Fig. 7, this time, we choose micro measures to focus more on the performance of samples. Figure 8 present the micro-averaged AUC score with respects to the number of training samples. It can be found that, for both algorithms, the performance increases significantly before the number of testing samples is equal to training samples from beginning, then with the gradual increase of the training set size, it grows with a slow rate until the end.

Looking closer at each graphs in Figs. 5, 6, 7 and 8, these results indicate that the predictive performance of these two ensemble algorithms becomes better with gradual increase of the training set size, whereas slight differences attributed to the variation of the intrinsic characteristics of each algorithms. For average precision and Hamming loss, their approximately invariable changing speed verifies the reasonable and satisfactory prediction quality indirectly. And the variation of macro precision, macro recall and micro averaged AUC suggest that one can get reasonable predictive performance of annotating new images when the number of testing samples is at least equal with training samples if one wants to cut the cost and manpower. Otherwise, the bigger training set size is, the better performance one could get.

5 Conclusions and Future Work

In this paper, we have proposed a new approach to supplement the description documents, where we cast the problem as an instance of multi-label learning. To achieve this, a product data crawled from several big online shopping webs is present. We develop a pre-processing procedure which can achieve multi-source data acquisition, fusion, “cluttered ” images removal and duplicate removal. This procedure can be easily generalized to add more categories to augment or update the data.

Based on this data, an application is developed to accomplish automatic and semi-manual annotation task with a set of basic labels which are picked from attribute data of the product and the elements of industrial design. To evaluate the quality of annotation labels, we have considered an extensive set of experiments, employing state-of-the-art multi-label learning algorithms under diverse and challenging scenarios. Experimental results suggest that almost all the features based on these two ensemble approaches could achieve quite high score, verifying the reasonable and satisfactory quality indirectly. In addition to the annotation work, the classification performance with respect to the number of training samples is further examined.

Multi-label learning algorithms trained on part of this data can also be employed to annotate newly acquired product images to generate corresponding description documents, greatly cutting the cost and manpower. Experimental results show that one can get satisfying performance of annotating new images when the number of testing samples is at least equal with training samples.

More subjective elements of industrial design will be done in future annotation which can better semantically describe the design proposal. Due to the difference for multi-label annotation between the product design field and traditional object understanding field, robust multi-label algorithms that suit for subjective labels in the product design field will be the main point in future work.

Acknowledgements This work is supported by the National Natural Science Foundation of China (No. 61772186).

References

1. Aksac, A., Ozyer, T., & Alhajj, R. (2017). Complex networks driven salient region detection based on superpixel segmentation. *Pattern Recognition*, *66*, 268–279.
2. Boutell, M. R., Luo, J., Shen, X., & Brown, C. M. (2004). Learning multi-label scene classification. *Pattern Recognition*, *37*(9), 1757–1771.
3. Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.
4. Cheng, M., Mitra, N. J., Huang, X., Torr, P. H. S., & Hu, S. (2011, June). Global contrast based salient region detection. In *IEEE conference on computer vision and pattern recognition* (pp. 409–416).
5. Chua, T., Tang, J., Hong, R., Li, H., Luo, Z., & Zheng, Y. (2009). NUS-WIDE: A real-world web image database from national university of singapore. *ACM International Conference on Image and Video Retrieval*, *48*, 1–9.
6. Clare, A., & King, R. D. (2002). Knowledge discovery in multi-label phenotype data. *Lecture Notes in Computer Science*, *2168*, 42–53.
7. Dimitrovski, I., Kocev, D., Loskovska, S., & Deroski, S. (2014). Fast and efficient visual codebook construction for multi-label annotation using predictive clustering trees. *Pattern Recognition Letters*, *38*(3), 38–45.
8. Dollar, P., Wojek, C., Schiele, B., & Perona, P. (2012). Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(4), 743–761.
9. Everingham, M., Gool, L. V., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, *88*(2), 303–338.
10. Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3), 75–174.
11. Gibaja, E., & Ventura, S. (2014). Multilabel learning: A review of the state of the art and ongoing research. *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, *4*(6), 411–444.
12. He, K., Zhang, X., Ren, S., & Sun, J. (2016, June). Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778).
13. Hou, Q., Cheng, M., Hu, X., Borji, A., Tu, Z., & Torr, P. H. S. (2019). Deeply supervised salient object detection with short connections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(4), 815–828.
14. Ioffe, S., & Szegedy, C. (2015, February). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448–456).
15. Jia, D., Russakovsky, O., Krause, J., Bernstein, M. S., Berg, A., & Li, F. F. (2014, April). Scalable multi-label annotation. In *SIGCHI conference on human factors in computing systems* (pp. 3099–3102).
16. Jia, X., Sun, F., Li, H., Cao, Y., & Zhang, X. (2017). Image multi-label annotation based on supervised nonnegative matrix factorization with new matching measurement. *Neurocomputing*, *219*, 518–525.
17. Jiang, M., Pan, Z., & Li, N. (2017). Multi-label text categorization using L21-norm minimization extreme learning machine. *Neurocomputing*, *261*, 4–10.

18. Karalas, K., Tsagkatakis, G., Zervakis, M., & Tsakalides, P. (2016). Land classification using remotely sensed data: Going multilabel. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(6), 3548–3563.
19. Li, J., Rao, Y., Jin, F., Chen, H., & Xiang, X. (2016). Multi-label maximum entropy model for social emotion classification over short text. *Neurocomputing*, *210*, 247–256.
20. Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014, April). Microsoft COCO: Common objects in context. In *European conference on computer vision* (pp. 740–755).
21. Luaces, O., Díez, J., Barranquero, J., del Coz, J. J., & Bahamonde, A. (2012). Binary relevance efficacy for multilabel classification. *Progress in Artificial Intelligence*, *1*(4), 303–313.
22. Madjarov, G., Kocev, D., Gjorgjevikj, D., & Deroski, S. (2012). An extensive experimental comparison of methods for multi-label learning. *Pattern Recognition*, *45*(9), 3084–3104.
23. Mas, J. A. D., & Marzal, J. A. (2016). Single users' affective responses models for product form design. *International Journal of Industrial Ergonomics*, *53*, 102–114.
24. Nowak, S., Lukashevich, H., & Dunker, P. (2010, January). Performance measures for multilabel evaluation: A case study in the area of image classification. In *International conference on multimedia information retrieval* (pp. 35–44)
25. Qi, Y., Zhang, G., & Li, Y. (2018). An auto-segmentation algorithm for multi-label image based on graph cut. *Sensing and Imaging*, *19*(1), 13–26.
26. Read, J., Pfahringer, B., Holmes, G., & Frank, E. (2011). Classifier chains for multi-label classification. *Machine Learning*, *85*(3), 333–359.
27. Rokach, L., Schclar, A., & Itach, E. (2014). Ensemble methods for multi-label classification. *Expert Systems with Applications*, *41*(16), 7507–7523.
28. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*(3), 211–252.
29. Santos, A. M., Canuto, A. M. P., & Neto, A. F. (2010, August). Evaluating classification methods applied to multi-label tasks in different domains. In *International conference on hybrid intelligent systems* (pp. 61–66)
30. Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *International conference on learning representations*.
31. Sivarajah, U., Kamal, M. M., Irani, Z., & Vishanth, W. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286.
32. Spolaôr, N., Lee, H. D., Takaki, W. S. R., & Wu, F. C. (2015). Feature selection for multi-label learning: A systematic literature review and some experimental evaluations. *International Journal of Computational Intelligence Systems*, *8*(2), 3–15.
33. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S. E., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015, June). Going deeper with convolutions. In *IEEE conference on computer vision and pattern recognition* (pp. 1–9).
34. Szymański, P., & Kajdanowicz, T. (2017, February). A scikit-based Python environment for performing multi-label classification. ArXiv e-prints.
35. Tomar, D., & Agarwal, S. (2015, October). Multi-label classifier for emotion recognition from music. In *International conference on advanced computing, networking and informatics* (pp. 111–123).
36. Tsoumakas, G., Katakis, I., & Vlahavas, I. (2011). Random k-labelsets for multilabel classification. *IEEE Transactions on Knowledge and Data Engineering*, *23*(7), 1079–1089.
37. Tsoumakas, G., Katakis, I., & Vlahavas, I. (2010, July). Mining multi-label data. In *Data mining and knowledge discovery handbook* (pp. 667–685).
38. Ukil, A. (2002). Support vector machine. *Computer Science*, *1*(4), 1–28.
39. Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010, June). SUN database: large-scale scene recognition from abbey to zoo. In *IEEE conference on computer vision and pattern recognition* (pp. 3485–3492).
40. Yang, Y. (1999). An evaluation of statistical approaches to text categorization. *Information Retrieval*, *1*(1), 69–90.
41. You, M., Liu, J., Li, G., & Chen, Y. (2012). Embedded feature selection for multi-label classification of music emotions. *International Journal of Computational Intelligence Systems*, *5*(4), 668–678.
42. Zhang, K., Wang, J., Hua, B., & Lu, L. (2013, October). DHash: A cache-friendly TCP lookup algorithm for fast network processing. In *38th annual IEEE conference on local computer networks* (pp. 484–491).

43. Zhang, M., & Zhou, Z. (2014). A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, *26*(8), 1819–1837.
44. Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, *237*, 350–361.
45. Zou, W., Liu, Z., Kpalma, K., Ronsin, J., Zhao, Y., & Komodakis, N. (2015). Unsupervised joint salient region detection and object segmentation. *IEEE Transactions on Image Processing*, *24*(11), 3858–3873.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Yong Dai[1,3] · Yi Li[2] · Li-Jun Liu[2]

Yong Dai
chd-dy@foxmail.com

Li-Jun Liu
sallyliu@hnu.edu.cn

[1] School of Electrical and Information Engineering, Hunan University, Changsha, Hunan Province, China

[2] School of Design, Hunan University, Changsha, Hunan Province, China

[3] Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Changsha, Hunan Province, China