Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8818616, 9 pages
https://doi.org/10.1155/2020/8818616

Research Article

Image Annotation via Reconstitution Graph Learning Model

Shi Chen,[1] **Meng Wang**,[2] **and Xuan Chen**[3]

[1]*Emporia State University, Kansas, USA*
[2]*Shandong Port Group Co., Ltd., Shandong, China*
[3]*Dalian University of Technology, Dalian, China*

Correspondence should be addressed to Meng Wang; dlutwangmeng@163.com

Received 20 September 2020; Revised 13 November 2020; Accepted 28 November 2020; Published 14 December 2020

Academic Editor: Xiaojie Wang

With great developments of computing technologies and data mining methods, image annotation has attracted much attraction in smart agriculture. However, the semantic gap between labels and images poses great challenges on image annotation in agriculture, due to the label imbalance and difficulties in understanding obscure relationships of images and labels. In this paper, an image annotation method based on graph learning is proposed to accurately annotate images. Specifically, inspired by nearest neighbors, the semantic neighbor graph is introduced to generate preannotation, balancing unbalanced labels. Then, the correlations between labels and images are modeled by the random dot product graph, to deeply mine semantics. Finally, we perform experiments on two image sets. The experimental results show that our method is much better than the previous method, which verifies the effectiveness of our model and the proposed method.

1. Introduction

With great developments of computing technologies and data mining methods, smart agriculture has attracted much attraction since it can greatly increase crop yields by effectively recommending methods to control pests [1, 2]. For example, the internet of vehicles with task scheduling [3, 4] can help formers to harvest crops automatically, and content-based crop image retrieval can help producers to keep track of plant growth in real time, which contributes to developing disease control and production plans. Meanwhile, with the technological advancement, the form of crop monitoring is also undergoing tremendous changes, posing great challenges to the current machine learning-based methods [5–9], due to the collected data that are of high volume, high velocity, high value, and high variety [10, 11]. Thus, to mine patterns of data in smart agriculture requires novel methods.

Image annotation, as a typical method for images analysis in agricultural big data, predicts labels for a given image, which can well match the image content [12]. In recent years, a large number of researchers have done extensive research on image annotation [13, 14]. For example, to reduce the semantic gap between visual features and text features, some researchers have proposed the generative model, which models image annotation as a joint likelihood distribution between images and labels. Nevertheless, the generative method only uses the image-label correlations, ignoring the relation over images. To use the relation over images, the discriminant model is proposed, focusing on finding the difference between images. Typically, this method trains a classifier to predict image labels, but the balance of sample labels has a large impact on the model performance. At the same time, some researchers proposed a graph model that utilizes all the data to build the intrinsic structure of unlabeled images and annotated images. Also, the nearest neighbor model is used to construct the label propagation graph, based on the theory that similar images share common labels [15, 16]. However, this method pays too much attention to the correlation between images, ignoring the image differences.

To solve those problems, a nearest neighbor graph model is proposed in this paper, which combining superiorities of graph and K nearest theories. Specifically, the semantic neighbors of test image under each label are firstly searched to the semantic neighbor graph. Then, a preannotation score

is obtained by graph learning of the semantic neighbor graph, considering relationships between images. The preannotation of the semantic neighbor graph can effectively solve the label imbalance problem, increasing the annotation probability of the rare labels and suppressing the high-frequency labels.

Next, the relationships between labels are used to improve the accuracy of the image annotation. The previous work was simply to calculate the cooccurrence probability between labels without considering the imbalance of cooccurrence between labels. For example, "Sea" and "Ship" are likely to appear in the same picture, and the two labels are strongly related. However, the possibility of "sea" in "ship" images may be greater than that of "ship" in "sea" images. This is because the "sea" is associated with more things, such as "fish" and "coral." To solve this imbalance of labels, the random dot product graph is used to mine the deep associations of labels. After that, visual differences that lead to lower similarity between similar images are used to further improve the performance of the proposed method. Finally, the naive Bayes nearest neighbor (NBNN) classifier is used to establish a joint likelihood between images and labels because of its simplicity and efficiency. Finally, the proposed method is conducted on Corel 5K and IAPR TC12. And results show that the proposed method has obvious improvement in terms of label recall. The main contributions of this paper are as follows:

(i) To effectively solves the label imbalance problem, the semantic neighbor graph learning is proposed to generate preannotation based on the nearest neighbor where all the labels are included in the initial label candidate

(ii) To mine the deep associations of labels, the random dot product graph is proposed, balancing the distributions of cooccurrence of paired labels

The remaining content structure of the thesis is as follows: in Section 2, we introduce the related work of image annotation. Then, in Section 3, we present our image annotation framework and concrete implementation of the framework. The datasets, experimental, settings, and results are illustrated in Section 4. The paper is concluded in Section 5.

2. Related Work

Image annotation has been a research hotspot which attracts increasing attention. Many fields are related to it, and they can benefit from the progress of each other. For example, the internet of vehicles [17, 18] can provide a lot of images to be annotated, and the better annotated images can be used to train the distinguishing model for better driving vehicles. Thus, a large number of researchers have introduced many kinds of methods to image annotation in recent years. They can be divided into four classes: the generating model, discriminating model, graph learning model, and nearest neighbor model.

2.1. Generating Model. To solve the problem in image annotation, some scholars proposed the mixture model, which is one of the generating model. For example, Jeon et al. proposed a cross-media relevance model (CMRM) [19]. In this method, image is segmented into several blobs, which can be clustered. Then, they calculate the probability between words and images by establishing maximum likelihood estimation. However, this method is affected by clustering of the image feature. Therefore, a continuous relevance model (CRM) [20] was proposed by Lavrenko et al., which used a continuous image feature. The method calculates the probability of labeling the word using polynomial distribution. But this method needs to store a large kernel matrix, resulting in a computational burden.

In order to solve the hybrid model's "visual ambiguity" problem, that visual similarities do not mean semantic similarities, researchers proposed the topic model. The topic model can be thought as a hybrid model with a particular topic used to portray the relationship between the image and the label. For example, Barnard et al. proposed a method with modeling multimodal cooccurrence [21]. This method imports several topic variables and attempts to find the relation between labels and visual features through probability. But this method is affected by model initialization. Blei et al. presented the LDA method [22], which used the Dirichlet distribution in the stage of choosing topics and words. However, the topic model is complex and has too many parameters. Thus, it is not suitable for large-scale datasets.

2.2. Discriminating Model. To solve the problem of the generating model, some researchers proposed the discriminating model. The discriminating model uses multilabel classification to solve the problem of image annotation. This method trains a classifier for each label, then determines which label the image belongs to by the classifier. For instance, Carneiro et al. proposed SML [23], which established a relationship between semantic labels and semantic classes. This method does not need to segment the image in advance, but it requires a high balance of classes and does not consider the relationship between labels. Sun et al. [24] used sparse factor representation to come up with sparse structure based on label dependency for weakening the negative effect caused by the unbalance of labels. But this method does not consider the potential relationship between images with labels and the lack of high-quality image dataset.

2.3. Graph-Based Learning Model. To address the issue of insufficient labeled images, some investigators put forward the graph learning model. The graph learning model is a semisupervised learning model, which uses labeled and unlabeled images to create the graph, then uses the Laplacian matrix for transferring labels. Liu et al. proposed the nearest spanning chain (NSC) [25]. In this method, they use a graph algorithm to transfer labels, but they do not take into account the relationship between images and labels. So Su and Xue proposed GLKNN [26]. In the stage of initializing graph weights, the cooccurrence relationship between labels is considered. However, they discount that the cooccurrence relationship is unbalance. This graph model only considers

Framework of the proposed method.
Input: images
Output: predicted labels
1: Find the best nearest neighbor images by improving the nearest neighbors.
2: Construct a similar matrix W through $W_{ij} = \exp[-\mathrm{DIS}(x_i, x_j)/\sigma^2]$.
3: Mine the deep relationship between images, using random dot product graph (RDPG) for refactoring, $P_X(G) = \prod_{i \neq j}(x_i \cdot x_j)^{a_{ij}} (1 - x_i \cdot x_j)^{1-a_{ij}}$.
4: Iterate to convergence through $R^*(t+1) = \alpha I \cdot R(t) + (1-\alpha)Y$.
5: Build a semantic matrix through $P(v_m \mid v_i) = \mathrm{sum}(m, i)/\mathrm{sum}(i)$
6: Consider the effect of the association between labels on the results of the annotations, $R'(t+1) = \alpha I \cdot R * (t) + (1-\alpha)P$.
7: Consider the relationship between images and labels, $\mathrm{Dis}(M, i) = \log(1/n)\sum_{K \in N(M.i)} K(d^M, d^K)$.
8: Return the final score of the label, $\mathrm{Score}(M, i) = \sigma R_{M,i} + \omega R^*_{M,i} + \xi \mathrm{Dis}(M, i)$.

Algorithm 1.

visual features and has no regard for problem of "visual ambiguity." Meanwhile, in the condition of a big image dataset, this model has high time complexity and poor annotation performance.

2.4. Nearest Neighbor Model. Because the nearest neighbor model performs better under big data conditions, this method has attracted more and more researchers. This model transfers the image annotation problem into the image retrieval problem. First, this method needs to search images which are highly similar to unlabeled images, then labels unlabeled images by means of label transmission. For example, Guillaumin et al. proposed a method based on weighted KNN called TagProp [27]. In this case, the labeled probability of lower frequency labels is increased and the labeled probability of higher frequency labels is suppressed. And Verma and Jawahar put forward 2PKNN [28], in which image distance metric learning was used. They adjust the weights of different visual features in order to make the relationship between visual features more consistent with the relationship between image semantics. In CCAKNN [29], the aim is to get the image subset of each semantic label. They map two features to the same subspace and model the visual features by using the Bayesian probability model. However, the nearest neighbor model only uses the similarity between images and ignores the difference between the image samples.

3. Our Approach

A new image annotation framework is proposed on the basis of graph learning, which is composed of three steps. First, we propose the nearest neighbor graph based on the principle that similar images share labels, to obtain preannotation results. Next, the association between labels is used to improve the accuracy of image annotation by the random dot product graph, which deeply mines the internal association of labels to increase probabilities of labeling weak labels. Finally, the naive Bayes nearest neighbor classifier is used to calculate the distance between images and labels. The main process of the proposed method is shown in Algorithm 1:

3.1. RPDG-Based Image Graph for Image Annotation. Let $X = \{x_1, x_2, \cdots, x_n\}$ be a collection of n images, $V = \{v_1, \cdots, v_l\}$ be a set of labels, and the training set be denoted by $T = \{(x_1, y_1), \cdots, (x_n, y_n)\}$, which is composed of each marked image x_i and corresponding label set y_i which is presented as a binary vector. For example, if the nth image is labeled by mth label, $y_n(m) = 1$; otherwise, $y_n(m) = 0$. To solve the problem of label imbalance, the nearest neighbor graph is constructed based on the neighbor image sets.

For a given image M to be labeled, its neighbor image set $\mathrm{Nei}(M)$ constitutes a set $S(M)$. We select a set of k nearest neighbor images to form a set T for each label based on the visual distance of images. The main idea of this method is that similar images have a high probability of passing labels. The traditional approach finds the semantic nearest neighbor by using the weighted multiple vision distance, without considering the probability that two images are neighbors to each other is different. As a result, the nearest neighbors of unlabeled images would have some noise images, which brings noise labeled and decreases the image annotation accuracy. Due to the complex distribution of visual features in images, some images in the image dataset have a higher probability of being selected as neighbor images, some images are less likely to be selected, and others may not even be selected. But in practice, the nearest neighbor relationship between images is not symmetric. For example, the image M is a nearest neighbor image of the image N, but the image N is not a nearest neighbor image of the image M, which degrades the accuracy of the conventional methods in selecting nearest neighbor images. Therefore, we propose a novel way to select the nearest neighbor images.

We propose a novel method based on the common neighbor images. We use this improved method to select the nearest neighbors of the test image, reducing the noise labels. Our method first sorts images of each label according to the visual distance and selects the first $2k$ images. Then, our method selects the nearest neighbors for each of these 2 k images. Sorting according to the number of their common neighbor images, the top K images are selected as the neighbor images of image M. The nearest neighbor images selected in this way are more consistent with the image similarity

under actual conditions. And the number of images which are related to the test image in semantic is also increased. As a result, the possibility of introducing a noise image is reduced and the accuracy of the annotation improves.

We assume a simple graph $G = (V, E)$, where V is the vertex set representing images in $S(M)$. The edge set is denoted as E representing a relationship between two images. The weight W of the edge is the similarity of two images.

The principle of the graph-based learning method is semisupervised learning. This method uses the image features and annotation information of the training data. Then, it iterates the similarity matrix of the training data and passes the appropriate semantic label from the labeled images to the unlabeled images based on this similarity, which is a preliminary result of the first step.

The detail of this method is as follows:

(Step 1) Construct a similar matrix $W^{k\times k}$ of $S(M)$ set as

$$W_{ij} = \exp\left[-\frac{\mathrm{DIS}(x_i, x_j)}{\sigma^2}\right], \tag{1}$$

where DIS() is a measure of distance. And W_{ii}=0, because there is no self-loop in the graph

(Step 2) Symmetrically normalize W by

$$I = D^{1/2} W D^{1/2}, \tag{2}$$

where D is a diagonal matrix and $D_{ii} = \sum_{j=1}^{l} W_{ij}$

(Step 3) Iterate according to the Eq. (3) until convergence

$$R(t+1) = \alpha I \cdot R(t) + (1-\alpha)Y, \quad R(0) = Y, \tag{3}$$

where t is the number of iteration until convergence and α is the propagation parameter

(Step 4) Label the unlabeled images according to the convergence matrix R^*

Through the above steps, we finally get the tag score and ranking. On the basis of the above discussion, there are two key parts of the graph-based learning method: a similarity graph (I) and an initial state representation (L). I describes the similarity between the test image and its nearest neighbor images, which provides a basis for the label transmission.

Thus, the construction of a similarity graph (I) is very important. In constructing a graph in the traditional graph-based image annotation methods, the weight of the edge between the vertices (images) directly uses the visual distance. However, because of the existence of the "visual ambiguity," this method may ignore the hidden relationships of the images. So different from the previous work, we use the random dot product graph to discover hidden relationships.

The random dot product graph is a point-edge random graph model. For each node $v_i, i = 1, \cdots, n$ in the node set V, a d-dimensional vector x_i is randomly and uniformly selected from the d-dimensional unit space as the assignment of v_i. The probability of the edge between each pair of nodes v_i, v_j is

$$p_{ij} = f(x_i \cdot x_j). \tag{4}$$

This probability is used for generating a random dot product graph as the assignment $X = [x_1, x_2, \cdots, x_n]_{d\times n}$.

The two main properties of random dot product graph are as follows:

Property 1. Clustering: the edges of random dot product graph appear with incompletely equal probability, with obvious clustering characteristics.

Property 2. Transitivity: if two nodes have strong connections with the third node at the same time, then the two nodes should also have a great correlation directly. Conversely, if two nodes have no other associated third node, then the probability that the two nodes are related should be small.

Each edge in the random graph appears randomly and independently. According to the Bernoulli distribution, the random dot product graph $G_x(V, E)$ generates the edge set E according to the probability p_{ij} to obtain an observation graph. If the observation graph $G = (V, E)$ is an undirected weighted graph and its adjacency matrix is $A = (a_{ij})_{n\times n}$, $a_{ij} \in [0, 1]$, then

$$P_X(G) = \prod_{i\neq j} (x_i \cdot x_j)^{a_{ij}} (1 - x_i \cdot x_j)^{1-a_{ij}}. \tag{5}$$

Its log likelihood function is

$$L_X(G) = \sum_{i\neq j} a_{ij} \ln (x_i \cdot x_j) + (1 - a_{ij}) \ln (1 - x_i \cdot x_j). \tag{6}$$

In the observation, the probability of the edge reflects the correlation between the nodes. It can be seen from Equation (6) that when $L_X(G)$ is maximum, the probability of the edge corresponds to the weight as much as possible. According to the principle of duality, we have

$$\max L_X(G) = \min F_Z(X), \tag{7}$$

where $F_Z(X) = \sum_{i\neq j} (x_i \cdot x_j - a_{ij})^2$.

Therefore, the objective function is expressed as

$$\min F_Z(X) = \min \sum_{i\neq j} (x_i \cdot x_j - a_{ij})^2 \tag{8}$$

where $X = [x_1, x_2, \cdots, x_n]$ is a random assignment of n nodes, the probability of the edge is the inner product form, and

Random dot product method for simple graphs.
Input: the weight matrix W of the image data graph.
Output: the weight matrix of random point product.
1: Take an all-zero matrix D.
2: Find spectral decomposition of $W + \text{diag}\ (D)$.
3: U is a matrix of d largest eigenvectors, $U \in \Re^{n\times d}$. $\tilde{\Lambda}$ is a $d \times d$ diagonal matrix composed of d largest eigenvalues, where each negative eigenvalue is changed to 0.
4: $X = \sqrt{U}\tilde{\Lambda}$, $D = \text{diag}\ (X'X)$
5: Return 2 until D converges.
6: Calculate $L_X(G)$, return 1 until converges. T is the edge probability matrix after random reconstruction, where $T = XX'$.

Algorithm 2.

the right side of Equation (8) is the Frobenius norm of the matrix, so it can be written as $A \approx X^T X$.

Based on the above principles, we have the following algorithm.

Based on the above method, for given nodes i and j, the $W_{ij}{}'$ weighted distance is expressed as

$$W_{ij}{}' = W_{ij} + \omega T_{ij}. \tag{9}$$

The random dot product graph improves the weight of the similar matrix. With the improvement of the nearest neighbor graph, we pay more attention to the internal relations between images. By this method, the weak label problem can be effectively solved.

3.2. Word-Based Graph Learning. The frequencies of the labels in the image dataset are different. The low-frequency labels are easily ignored during the annotating process, which leads to the accuracy decrease of the annotation. In the previous work, people usually used the semantic symbiosis between labels to solve this problem. However, there is a cooccurrence imbalance between the labels, which makes it impossible to significantly improve the label effect of the low-frequency label. By the transitive nature of the random dot product graph, we reconstruct the association graph of the label words and find the inherent hidden relationship between the labels. The random dot product graph can obtain the relationship between any annotated words. The probability of common semantic relations is large, and the probability of uncommon semantic relations is small, which is consistent with the real semantic relationship.

In the label set $V = \{v_1, \cdots, v_l\}$, we record the probability of label v_i to label v_m denoted by $P(v_m \mid v_i)$,

$$P(v_m \mid v_i) = \begin{cases} \dfrac{\text{sum}(m,i)}{\text{sum}(i)}, & m \neq i, \\ 1, & m = i, \end{cases} \tag{10}$$

where $\text{sum}(m, i)$ represents the number of cooccurrences between labels v_i and v_m. In this paper, we abbreviate $P(v_m \mid v_i)$ to P_{im}. Because of semantic cooccurrence imbalance, P_{im} is not equal to P_{mi}.

We first get the transfer matrix between the labels according to Equation (10). P is reconstructed by random dot product to obtain P'. Bringing transfer matrix and the matrix $R*$ obtained on the basis of graph learning into Equation (3), we iterate to get result R'.

3.3. Image to Word Relation. This relationship can be regarded as the possibility of having an image to produce a label. In most cases, the relationship can be estimated on a training set by some hypothetical distribution. In many methods, the image is clustered and divided into several "blob," with each "blob" corresponding to a label word. However, in the process of clustering, problems will be caused due to that the underlying features are similar, but the actual contents are different, which makes the blob itself wrong. In this paper, the method used to calculate the image to word distance is the naive Bayes nearest neighbor (NBNN) classifier [30] for image classification. This method is simple and has good performance. At the same time, it calculates the association between the whole image and the annotated label, avoiding the wrong correspondence between the "blob" and the annotated label.

The features of the image are recorded as f, and $N(M, i)$ represents a collection of $\text{Nei}(M)$ annotated as label vi. The definition of image to word distance is

$$\text{Dis}(M, i) = \log \frac{1}{n} \sum_{k\in N(M.i)} K\left(f^M, f^k\right), \tag{11}$$

where n is the figure for images in $\text{Nei}(I, k)$. $K()$ is the Gaussian kernel function:

$$K\left(f^M, f^k\right) = \exp\left(-\frac{1}{2\sigma^2}\left\|f^M - f^k\right\|^2\right). \tag{12}$$

3.4. Combination of Three Scores. Finally, we combine the two scores based on the graph learning with the score of the image-to-label distance to get the final score, which is the basis for the final labels.

$$\text{Score}(M, i) = \sigma R_{M,i} + \omega R^*_{M,i} + \xi \text{Dis}(M, i), \tag{13}$$

where $R'_{M,i}$ is a score based on the association between images and $R^*{}_{M,i}$ is the probability that the image M is labeled with the label v_i based on an association between labels. In addition, $\sigma + \omega + \xi = 1$.

4. Experiment

In this section, we introduce two datasets used in the experiment and the extraction of features of two datasets. Also, the evaluation indicators of the image annotation methods are given.

4.1. Datasets. During the experiment, we used two datasets. Table 1 shows the statistics of these datasets.

Corel 5K [31]. This dataset contains 4,500 training images and 499 test images. It is divided into 50 themes, each with 100 images except the last. The dataset contains 260 labels. Each image is manually labeled with 1-5 different labels, and the average is 3.4.

IAPR TC12 [32]. This dataset contains 19,627 images, where 17,665 are training images and 1962 are test images. This dataset contains a total of 291 tags, and each image in the dataset is averaged as 5.836 tags.

4.2. Feature. The first step in our approach is to extract features, which is a very important part. Feature extraction has a profound impact on the performance of image annotation systems. Recently, CNN has been widely applied to feature extraction of images. Compared with using 15 handcrafted features, it is not necessary to use metric learning to determine the optimal weight of each feature, so it is easier to determine the parameters. We use CNN to extract individual features instead of handcrafted features, which can effectively reduce the number of features and improve system accuracy.

4.3. Evaluation Metrics. In our experiments, we use the same evaluation method as [33] to effectively evaluate and compare our method with the previous methods. In our approach, we give each image five labels. Then, we calculate the labeling precision and recall for each label in each image of the test set. Suppose that a label v_k marking n_1 images in the ground truth, and the number of images marked as v_k during the test is n_2, in which the correct number of marks is recorded as n_3. The method of calculating the precision of the label v_k is $p = n_3/n_2$ and recall of the label v_k is $r = n_3/n_1$. These values are obtained by calculating each label, and then, the mean value is calculated to get the average precision P and the average recall R. Define that F1 is the score for combining P and R, $\mathrm{F1} = 2PR/(P+R)$. And define that $N+$ represents the number of tags that have been correctly tagged at least once, which indicates the ability of our proposed method to solve class imbalance and weak label problems.

5. Result

In this subsection, we describe the performance of the proposed method compared with the previously proposed methods. Table 2 gives the experimental results on the datasets Corel 5K and IAPR CT12. This table shows that our method outperforms the previous work. Among the Corel 5K, our accuracy is the second highest, and our tag recall number is the highest. Detailed results and analysis of the experiment will be presented in the following sections.

It is worth noting that we have selected several methods based on nearest neighbors as comparison methods. As shown in Table 2, our method performs better than JEC in all aspects. Compared with 2PKNN, our recall value and $N+$ value is also much higher on the Corel 5K dataset. And our RDPGKNN is superior to TagProp. The comparison with these methods shows that the graph learning method also has unique advantages in the field of image annotation and proves the validity and rationality of the label using the graph learning method for propagation.

Table 1: Details of the training set of the two datasets. The number of images and labels are given in the format mean/maximum.

Dataset	Corel 5K	IAPR TC12
# of img.	4999	19627
Vocab. size	260	291
Training img.	4500	17665
Testing img.	499	1962
Labels per img.	3.4/5	5.836/10.04
Img. per label	5.7/23	347.7/4999

We also compare RDPGKNN with graph-based learning algorithms, and the results show that our approach is generally better than previous work. Since most of the graph learning algorithms are applied to small vocabularies, there are few research methods on image annotation based on graph learning in Corel 5K and other datasets, so we mainly choose TGLM and GLKNN. In comparison with TGLM, the experimental results show that our method is obviously superior in Corel 5K. This shows the advantage of the nearest neighbor method, which effectively solves the label imbalance problem, so that each annotation word has the opportunity of being selected. At the same time, compared with GLKNN, our $N+$ has a significant improvement, because we consider label cooccurrence asymmetry. Using graph-based learning to calculate the label transition probability can maximize the selected probability of low-frequency tags, provide more appropriate weights for the transfer between tags, and improve the performance of the image tagging system.

On the IAPR CT12, our algorithm also has excellent performance. Compared with the previous work, the RDPGKNN method recalls the most labels. On this basis, our recall rate is second only to that of the CAAKNN method, and the recall rate is greatly improved on the premise that the accuracy does not drop too much. Compared to the GLKNN based on the graph, the recall rate of our method has also increased by 2%. This also confirms the need to consider the problem of cooccurrence imbalance between images. Figure 1 shows some examples of the annotation of our method on two datasets. Among them, we use the black mark to indicate the labels annotated in ground truth and annotated with RDPGKNN, and the red mark does not appear in the ground truth. It should be noted that some images in the dataset have fewer than five labels annotated in ground truth, but our method must label five labels.

After comparing with all methods, we find that our method effectively increased the value of $N+$. This shows that compared with the traditional methods, our method has strong performance in recall, and the other performance

Table 2: The performance of our proposed method is compared with the previous work on Corel 5K and IAPR TC12 datasets in detail. P: average precision; R: average recall; F: the combination of P and R; $N+$: number of labels with nonzero recall value.

Method	Corel 5K				IAPR TC12			
	P (%)	R (%)	F (%)	$N+$	P (%)	R (%)	F (%)	$N+$
CRM [20]	16	19	17	107				
MBRM [34]	24	24	24	122	24	23		223
SML [23]	23	29	26	137				
JEC [35]	27	32	29	139	29	28		250
TGLM [25]	25	29	27	131				
TagProp σSD [27]	28	35	31	145	41	30		259
TagProp ML [27]	31	37	34	146	48	25		227
TagProp σML [27]	33	42	37	160	46	35		266
KSVM-VT [33]	32	42	36	179	47	26		268
FastTag [36]	32	43	37	166	47	26		280
GLKNN [26]	36	47	41	184	41	36		282
2PKNN [28]	39	40	39	177	49	32		274
LDMKL [37]	29	44	35	179				
IDFRW [38]	38	49	43	185	49	31	38	275
CCAKNN [29]	41	43	42	185	41	40	41	278
RDPGKNN (this work)	40	45	40	195	40	38	38	283

Test image					
Ground truth	Sun, Water, Clouds, Birds	Forest, Cat, Tiger, Bengal	Leaf, Petals, Stems, Flowers	Wall, Cars, Tracks, Formula	Bear, Polar, Snow, Tundra
Predicted labels	Sun, Sea, Horizon, Waves, Land	Forest, Cat, Tiger, Bengal, Park	Leaf, Flowers, Petals, Blooms, Needles	Wall, Cars, Tracks, Formula, Plaza	Bear, Polar, Snow, Tundra, Cubs
Test image					
Ground truth	Ice, Frost, Frozen	Grass, Bear, Grizzly, Meadow	Sky, Jet, Plan	Sky, Train, Railroad, Locomotive	Coral, Ocean, Reefs
Predicted labels	Ice, Frost, Frozen, Branch, Relief	Grass, Bear, Grizzly, Meadow, Calf	Sky, Jet, Plan, F-16, Fly	Train, Flag, Railroad, Locomotive,Vehicle	Coral, Ocean, Reefs, Fan, Pots

Figure 1: The result of annotating some images in the Corel 5K dataset of our method. The figure shows the test image, the ground truth labels, and the predicted labels, where red indicates that the label is not present in the ground truth labels.

is almost unchanged. Also, the problem that some labels cannot be selected due to the unbalanced label co-occurrence phenomenon is solved.

6. Conclusion

In this paper, a reconstitution graph learning model is proposed to for image annotation in smart agriculture. To solve the weak label problem, a nearest neighbor graph learning model is proposed to get the prelabels. Meanwhile, for the cooccurrence imbalance between labels, the random dot product graph is used to explore the intrinsic links between labels. Many experiments on the Corel 5K and IAPR TC12 are conducted, and the result shows that the recall of our method is much larger than that of the previous graph-based learning methods. At the same time, our accuracy and recall rate are basically the same as the latest methods. In the future, we will force on the computational complexity

of the proposed method and the depth correlation between labels and images in the annotation process.

Data Availability

The datasets used in this paper are public datasets which can be accessed by the following websites: Corel 5K: https://rdrr.io/cran/mldr.datasets/man/corel5k.html and IAPR TC12: http://www-i6.informatik.rwth-aachen.de/imageclef/resources/iaprtc12.tgz;

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by "the Fundamental Research Funds for the Central Universities", No. DUT20LAB136.

References

[1] I. Mat, M. R. M. Kassim, A. N. Harun, and I. M. Yusoff, "Smart agriculture using internet of things," in *Proceedings of the 2018 IEEE Conference on Open Systems (ICOS)*, Langkawi, Malaysia, 2018.

[2] F. Bu and X. Wang, "A smart agriculture IoT system based on deep reinforcement learning," *Future Generation Computer Systems*, vol. 99, pp. 500–507, 2019.

[3] X. Wang, Z. Ning, S. Guo, and L. Wang, "Imitation learning enabled task scheduling for online vehicular edge computing," *IEEE Transactions on Mobile Computing*, p. 1, 2020.

[4] Z. Ning, P. Dong, X. Wang et al., "Partial computation offloading and adaptive task scheduling for 5G-enabled vehicular networks," *IEEE Transactions on Mobile Computing*, p. 1, 2020.

[5] P. Li, Z. Chen, L. T. Yang, J. Gao, Q. Zhang, and M. J. Deen, "An improved stacked auto-encoder for network traffic flow classification," *IEEE Network*, vol. 32, no. 6, pp. 22–27, 2018.

[6] P. Li, Z. Chen, L. T. Yang, Q. Zhang, and M. J. Deen, "Deep convolutional computation model for feature learning on big data in internet of things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 790–798, 2018.

[7] J. Gao, P. Li, Z. Chen, and J. Zhang, "A survey on deep learning for multimodal data fusion," *Neural Computation*, vol. 32, no. 5, pp. 829–864, 2020.

[8] X. Wang, Z. Ning, and S. Guo, "Multi-agent imitation learning for pervasive edge computing: a decentralized computation offloading algorithm," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 2, pp. 411–425, 2021.

[9] J. Gao, P. Li, and Z. Chen, "A canonical polyadic deep convolutional computation model for big data feature learning in internet of things," *Future Generation Computer Systems*, vol. 99, pp. 508–516, 2019.

[10] B. Rani, M. Kumari, K. Sobha, P. Kumari, J. Majhi, and S. Chakraborty, "Application of Big Data in Smart Agriculture," *SSRN Electronic Journal*, 2020.

[11] Z. Ning, P. Dong, X. Wang et al., "Mobile edge computing enabled 5G health monitoring for internet of medical things: a decentralized game theoretic approach," *IEEE Journal on Selected Areas in Communications*, 2020.

[12] Q. Cheng, Q. Zhang, P. Fu, C. Tu, and S. Li, "A survey and analysis on automatic image annotation," *Pattern Recognition*, vol. 79, pp. 242–259, 2018.

[13] Y. Sun and K. A. Loparo, "Context aware image annotation in active learning," 2020, http://arxiv.org/abs/2002.02775.

[14] Y. Niu, Z. Lu, J. R. Wen, T. Xiang, and S. F. Chang, "Multi-modal multi-scale deep learning for large-scale image annotation," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1720–1731, 2019.

[15] B. N. Tandel and U. Desai, "Various face annotation techniques: survey," in *Intelligent Communication Technologies and Virtual Mobile Networks*, pp. 94–102, Francis Xavier Engineering College, Tamil Nadu, Tirunelveli, India, 2019.

[16] M. Sangeetha, K. Anandakumar, and A. Bharathi, "Automatic image annotation and retrieval: a survey," *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 4, pp. 1143–1147, 2016.

[17] Z. Ning, K. Zhang, X. Wang, L. Guo, and R. Y. K. Kwok, "Intelligent edge computing in internet of vehicles: a joint computation offloading and caching solution," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–14, 2020.

[18] Z. Ning, R. Y. K. Kwok, K. Zhang et al., "Joint computing and caching in 5G-envisioned internet of vehicles: a deep reinforcement learning based traffic control system," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12, 2020.

[19] J. Jeon, V. Lavrenko, and R. Manmatha, "Automatic image annotation and retrieval using cross-media relevance models," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval - SIGIR '03*, pp. 119–126, Toronto, Canada, 2013, ACM.

[20] V. Lavrenko, R. Manmatha, and J. Jeon, "A model for learning the semantics of pictures," *Advances in Neural Information Processing Systems*, vol. 16, pp. 553–560, 2003.

[21] K. Barnard, P. Duygulu, D. Forsyth, N. D. Freitas, D. M. Blei, and M. I. Jordan, "Matching words and pictures," *Journal of Machine Learning Research*, vol. 3, pp. 1107–1135, 2003.

[22] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[23] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos, "Supervised learning of semantic classes for image annotation and retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 394–410, 2007.

[24] F. Sun, J. Tang, H. Li, G. J. Qi, and T. S. Huang, "Multi-label image categorization with sparse factor representation," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1028–1037, 2014.

[25] J. Liu, M. Li, Q. Liu, H. Lu, and S. Ma, "Image annotation via graph learning," *Pattern Recognition*, vol. 42, no. 2, pp. 218–228, 2009.

[26] F. Su and L. Xue, "Graph learning on K nearest neighbours for automatic image annotation," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pp. 403–410, Shanghai, China, 2015.

[27] M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "Tag Prop: discriminative metric learning in nearest neighbor models for image auto-annotation," in *2009 IEEE 12th International Conference on Computer Vision*, pp. 309–316, Kyoto, Japan, Oct. 2009.

[28] Y. Verma and C. V. Jawahar, "Image annotation using metric learning in semantic neighbourhoods," in *Proceedings of the*

12th European conference on Computer Vision, pp. 836–849, Springer, Berlin, Heidelberg, 2012.

[29] X. Wang, H. Ge, and L. Sun, "Image automatic annotation algorithm based on canonical correlation analytical subspace and k-nearest neighbor," *Journal of Ludong University(Natural Science Edition)*, vol. 32, no. 2, pp. 97–104, 2018.

[30] O. Boiman, E. Shechtman, and M. Irani, "In defense of nearest-neighbor based image classification," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA, June 2008.

[31] https://rdrr.io/cran/mldr.datasets/man/corel5k.html.

[32] http://www-i6.informatik.rwth-aachen.de/imageclef/resources/iaprtc12.tgz.

[33] Y. Verma and C. V. Jawahar, "Exploring SVM for image annotation in presence of confusing labels," in *British Machine Vision Conference 2013*, Bristol, UK, 2013.

[34] S. L. Feng, R. Manmatha, and V. Lavrenko, "Multiple Bernoulli relevance models for image and video annotation," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*, vol. 2, pp. 1002–1009, Washington, DC, USA, 2004.

[35] A. Makadia, V. Pavlovic, and S. Kumar, "A new baseline for image annotation," in *Proceedings of 10th European Conference on Computer Vision*, Marseille, France, 2008.

[36] M. Chen, A. Zheng, and K. Weinberger, "Fast image tagging," in *Proceedings of the 30th International Conference on Machine Learning, PMLR*, pp. 1274–1282, Atlanta, GA, USA, 2013.

[37] M. Jiu and H. Sahbi, "Nonlinear deep kernel learning for image annotation," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1820–1832, 2017.

[38] Z. Ning, G. Zhou, Z. Chen, and Q. Li, "Integration of image feature and word relevance: toward automatic image annotation in cyber-physical-social systems," *IEEE Access*, vol. 6, pp. 44190–44198, 2018.