Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1381-1386 (2020)

www.astesj.com

Special Issue on Multidisciplinary Sciences and Engineering

ASTESJ

ISSN: 2415-6698

Image Tag Recommendation based on Ranked Categorical Nearest Neighbors and Weighted Tag Features

Anupama D. Dondekar*, Balwant A. Sonkamble

Department of Computer Engineering, Pune Institute of Computer Technology, Pune, 411042, India

ARTICLE INFO

Article history:
Received: 16 October, 2020
Accepted: 24 November, 2020
Online: 16 December, 2020

Keywords:
Tag Recommendation
Categorical Nearest neighbor
Weighted Tag Frequency
Flickr

ABSTRACT

The huge number of images on the image sharing websites poses challenges for classification and retrieval of the images. On many image sharing websites, tags can be assigned by the users to an image that describes the contextual and visual description of an image. However, ambiguous or incorrect tags have appeared in frequent tags that affect the performance of an image retrieval system. Thus, assigning appropriate tags to the images plays a very important role in image retrieval and classification. In this paper, the ITR-WTF image tag recommendation method is proposed which explores tags from ranked nearest neighbors of each category. For a given input image, the method first determines the neighbors from training images of each category and ranks the neighbors according to the distance from the input image. In the second step, the weight is assigned to each tag based on the vote from each neighbor. Finally, the weighted tag frequency is determined to recommend appropriate tags to a given image. The experimentation is done on two datasets self-generated and NUS-WIDE. The results obtained using the proposed method ITR-WTF gives good results as compared with the existing methods of tag recommendation.

1. Introduction

The rapid development of advanced technology and the high usage of social media have created a large repository of images which poses many challenges for an image retrieval system. On many image sharing social websites, the users are uploading images for faster communication or to find the people with the same interest. The images on the image sharing social websites are associated with the tags which are assigned by the users. The tags describe the visual content of the images along with the context information such as the location and time the images are captured. These tags are used indexing during an image search. Thus, tags play a very important role in tag-based image retrieval systems and classification.

Many times, users assign tags that are imprecise and irrelevant to the image. According to the survey, only 50% of the Flickr tags describe the content of the images [1]. The presence of irrelevant tags affects the performance of the classification and image retrieval system. Hence, it is necessary to design an algorithm that assigns correct tags to the images. It improves the accuracy of the image retrieval system by suggesting proper tags for the images. The advantages of tag suggestions are: it reduces the cost of manual annotation of the images, spelling mistakes. Several studies have been done on tag recommendation using visual contents, tags and metadata. Still, the performance of tag recommendation is not satisfactory due to personalized tags as shown in Figure 1. The tags may not describe the content of the images and may be assigned as per the user perspective. The tags Denmark, 2011 does not describe visual content of an image as shown in Figure 1 (a).

(a) 

Tags:

Denmark, 2011, nature, D60

(b) 

Tags:

Fish, Canon 50mm, Canon EOS 5D Mark III

Figure 1: Tag based Images

*Corresponding Author: Anupama D. Dondekar, Email: agphakatkar@pict.edu

In this paper, the ITR-WTF image tag recommendation method is proposed which contributes in the following way:

- Obtained ranked visual neighbor of each category for each feature
- The tag score is calculated by combining weighted tf-idf value and vote from rank categorical nearest neighbor to recommend top k tags

The paper is organized as follows: existing methods for tag recommendation are described in section 2. Section 3 explains the ITR-WTF proposed method for image tag recommendation. Section 4 and 5 describe the metric used for evaluating the performance of the proposed method and dataset used for experimentation. Section 6 describes the experimental results. Finally, in section 7 the conclusion and future work is given.

2. Work Done

The methods for tag recommendation are categorized into three methods: classification, semantic and nearest neighbor based.

In classification-based methods, the features of the images are extracted and classified into a category using a classification algorithm. The classifier trained for multiple class result in a multi-class problem. Finally, the tags are recommended based on the category of an input image. In [2] the method was proposed labeling of the images. The images were segmented into regions and identified salient region. Based on particle swarm optimization, the SVDD trained to assign labels to the images by assigning more weights to the salient regions. In [3] the method was proposed using probability and weight based SVM classifier for annotation of images. Given an image with tag, the method identifies the related and unrelated images using majority of voting from SVM in [4]. The system named 'SheepDog' was developed to which identifies the suitable group for the inclusion of photos and suggest appropriate tags to the users on Flickr dataset in [5].

In the semantic method, the tags are recommended based on the joint distribution of image and tag features. The method was proposed in [6] for annotation of images using KCCA framework by constructing semantic space in which the correlation was built between image feature and tag features. In [7] an approach for retagging of social images with diverse semantics was presented. Both the relevance of a tag and the semantic compensation to the already determined tags was fused to determine the final tag list for a given image. The method was proposed in [8] to recommend tags for geotagged images using unified subspace which correlates the textual and visual features. The hyper-graph-based method was proposed for tag-based image retrieval using image features and tag features simultaneously [9].

The nearest neighbor-based tag recommendation methods are very popular due to its effectiveness. The model is based on the assumption that feature based similar images tend to have the same tags. Given an input image, the method determines k nearest neighbors by combining various features either by using early or late fusion and calculates the tag relevance score by collecting votes of a tag from its nearest neighbors. The advantages of the method are: it is scalable and model building is not needed as it makes an assumption based on training data. The tag recommendation method was proposed in [10] using random walk on bipartite graph constructed based on weighted user and image nearest neighbor. In [11], the Bayesian based image annotation model was proposed based on semantic nearest neighbors. The method was proposed in [12] for annotation of images using a variation of traditional kNN algorithm by defining matrix which shows the relationship between labels and images. In [13] the method was proposed in which the given an image the similar images were determined using k nearest neighbor and tag graph was created from tags of neighbors and clustered to assign label to an image. The personalized image tag recommendation method was proposed based on neighbor voting scheme by building tripartite graph to show relationship between user, tags and images in [14]. The VS-KNN method was proposed in [15] for image labeling by exploring image features and label features simultaneously as a maximum posterior estimation. Given an image with label l, the method was proposed which identify the images labeled with l using kNN and denote it by set S [16]. Finally, the labels were assigned to the images by calculating the similarity between the input image and set S. The tag relevance method was proposed by assigning weight to each neighbor based on distance using kNN [17]. The method was proposed in [18] to suggest tags for an image based on visual features and tag correlations using neighbor voting scheme. Photo tagging method was proposed in [19] using history of the users. The method finds the geographical, visual and time neighbors for a given image and recommends tags by counting accumulating votes for each tag from three types of neighbors. In [20], the method was proposed to suggest tags to the images with and without labels. The method first identifies a set of k images using feature-based k nearest neighbor. Finally, assign or recommends the tags by counting the difference between tag frequency count from the entire database and k neighbor. The image annotation algorithm in [21] identifies rank based and weight based nearest neighbor and suggests the label for an input image using a probabilistic model.

In this paper, the proposed method recommends tags based on the nearest neighbor method. Compared with the classification and semantic based method, the nearest neighbor method is popular due to its effectiveness and scalability. Also, it does not require any training.

However, the performance of the existing nearest neighbor-based tag recommendation methods depends on the number of k neighbors and may be affected due to the equal weight and voting irrespective of the class. This observation motivated us to develop a method which improves the performance of tag recommendation. The method first determines the weighted categorical nearest neighbors and suggest the tags based on the ranked categorical neighbor and weighted tag frequency.

3. Research Methodology

In this section, ITR-WTF proposed tag recommendation algorithm is described. The main objective of the proposed algorithm is to improve the accuracy of tag recommendation by identifying distance based nearest neighbor and rank the tags by combining image score and tag score.

3.1. ITR-WTF Tag Recommendation Method

The block diagram of the proposed method for tag recommendation is shown in Figure 2. The proposed method consists of three main modules: feature representation, classification and tag recommendation.

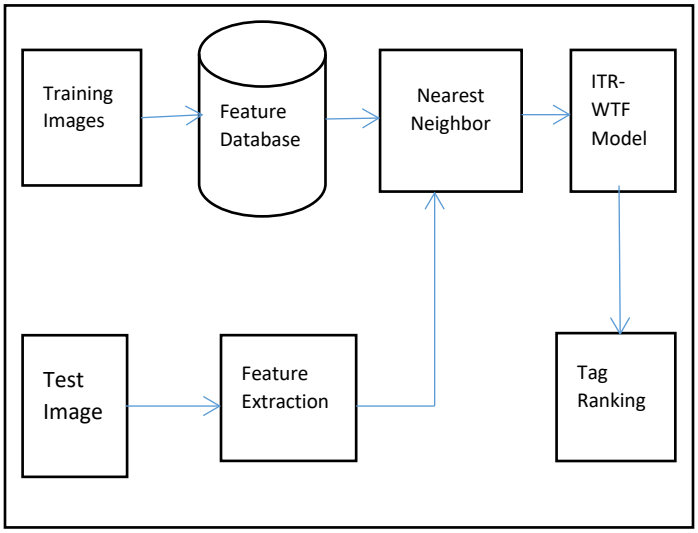

Figure 2: System Diagram of Proposed Method

3.1.1. Feature Representation:

Features play a very important role in image representation. Several researchers have worked on feature extraction methods for image retrieval using color, texture and shape features. Extracting an effective feature and represent them efficiently is very important.

Color is the most used feature in an image retrieval system. Using color features, a human can recognize most images and objects included in the images. Also, the color features are invariant to scaling, translation and rotation of an image. Another important feature is texture. Texture measures look for visual patterns in images and how they are spatially defined.

During the training phase, the features are extracted using color moment and wavelet packet transform [22, 23]. The first, second and third moments are extracted as color features in L*a*b* color space by segmenting an image into two by two sub-block along with a centralized sub-block of the same size resulting into 9 features for each sub-block. For texture features are extraction using wavelet packet transform, an image is decomposed into sub-bands up to level three using daubechies wavelet. The energy and standard deviation of each band are determined as texture features using eq.1 and 2.

$$M_i = \frac{1}{HT*WT} \sum_{h=1}^{HT} \sum_{w=1}^{WT} |Coeff_i(h,w)| \quad (1)$$

$$SD_i = \sqrt{\frac{1}{HT*WT} \sum_{h=1}^{HT} \sum_{w=1}^{WT} (|Coeff_i(h,w)| - M_i)^2} \quad (2)$$

where i = 1 to 4^L and L=3, $Coeff_i(h,w)$ represents the coefficient values of i^{th} decomposed sub-band image at level L, WT and HT are the width and height of the decomposed sub-band image $Coeff_i$.

Each feature has a different range of values. To avoid the influence of one feature due to variation in a range of values, the features are normalized between 0 and 1 range using min-max normalization method.

There are two methods used for combining features: early and late fusion. In early fusion, the individual features are combined before image similarity score calculation. In late fusion, the image similarity score determines for each feature and combine the individual score to calculate the final score. The late fusion method needs more cost of computation [24]. For this reason, the early fusion technique is used to determine the image similarity score.

3.1.2. Nearest neighbor:

After feature extraction, the nearest neighbor classifier determines the neighbors of a test image. In the nearest neighbor classifier, the neighbors for each class/category are determined as shown in Figure 3 and combined to form the final neighbors. The harmonic mean of the neighbors of each category is determined and finally predicts the category with the smallest mean [25].

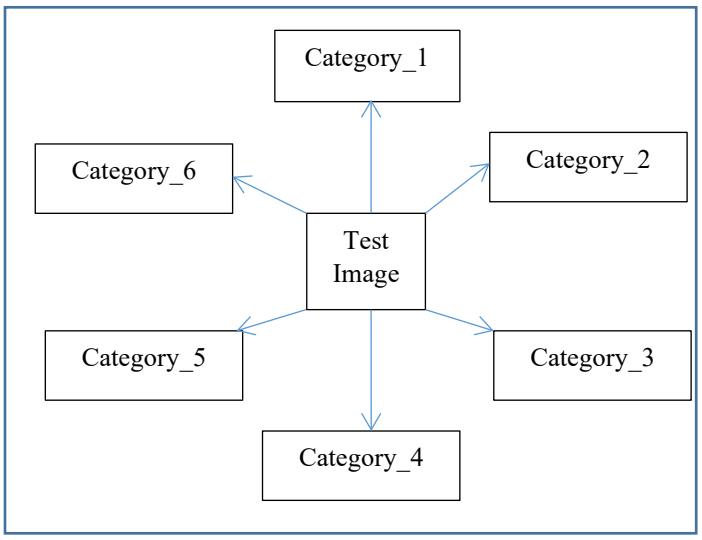

Figure 3: Nearest Neighbor from Each Category

To improve the effectiveness of feature-based neighbors, the weight is assigned to each neighbor so that nearby training image will get more weight and the training images which are farther away will be assigned less weight.

3.1.3. ITR-WTF Model:

Once the feature based neighbors are determined, the method assign an importance to each tag based on image similarity score as follow:

$$P(t) = \sum_{x_i \in NN} sim(I_q, x_i).\delta(x_i, t) \quad (3)$$

where $sim(I_q, x_i)$ denotes the importance of image x_i in predicting tag t according to their visual similarity; $\delta(x_i, t)$ is equal to one if tag is associated with an image x_i, otherwise $\delta(x_i, t)$ is equal to zero. Finally, the score of tag t is calculated as follow:

$$w - tfidf(t) = \left[1 + \log(N_t) * \log\left(\frac{NN}{N_t}\right) * P(t)\right] \quad (4)$$

where N_t represents the number of images associated with t and NN represents the total number of nearest neighbors.

3.2. Algorithm

Algorithm 1: The proposed ITR-WTF Algorithm

Input:

CV: Color feature vector of an input image
TV: Texture feature vector of an input image
$TFV_j = (y_i^j, C_i^j)_1^{NF_j}$: j[th] class texture feature vector of training images
$CFV_j = (x_i^j, C_i^j)_1^{NF_j}$: j[th] class color feature vector of training images
M = $C_1, C_2, \ldots\ldots C_M$: the number of class labels
$T_{mn} = \{T_{1n}, T_{2n}, T_{3n}, \ldots\ldots\ldots T_{mn}\}$: Tag associated with n[th] image

Output: Ranking of Tags

Step 1: Calculate the distance between CV and CFV_j using eq. (3). The set of m number of nearest neighbor for each class C_j is denoted as $CD_{mj}^{NF} = [x_{1j}^{NF}, x_{2j}^{NF}, \ldots\ldots, x_{mj}^{NF}]$

$$dist(CV, x_i^j) = \sqrt{\sum_{l=1}^{m}(CV_l - x_{il}^j)^2} \quad (5)$$

where $x_i^j \epsilon\, CFV_j$

Step 2: Calculate the distance between TV and TFV_j using eq. (4). The set of n number of nearest neighbor for each class C_j is denoted as $TD_{nj}^{NF} = [y_{1j}^{NF}, y_{2j}^{NF}, \ldots\ldots, y_{nj}^{NF}]$

$$dist(TV, y_i^j) = \sqrt{\sum_{l=1}^{n}(TV_l - y_{il}^j)^2} \quad (6)$$

where $y_i^j \epsilon\, TVF_j$

Step 3: Merge the nearest neighbors obtained using color and texture features

$NN = \{CD_{ij}^{NF} \cup TD_{ij}^{NF}\}$ where i = 1……M

Step 4: Assign an importance to each tag based on image similarity score as follows:

$$P(t) = \sum_{x_i \in NN} sim(I_q, x_i).\delta(x_i, t)$$

where

$$sim(I_q, x_i) = dist(I_q, x_i) * wt_j^i$$

$$wt_j^i = \frac{1}{j} \qquad i = 1, \ldots\ldots, m$$

Step 5: Calculate the w-tfidf of each tag as follows

$$w - tfidf(t) = \left[1 + \log(N_t) * \log\left(\frac{NN}{N_t}\right) * P(t)\right]$$

Step 6: Rank the tags according to the w- tfidf values and select top k values

4. Performance Metric

To evaluate the performance of proposed method, we use NDGC, Precision, Recall and F1-score.

Given an image with ranked tag list $T_1, T_2, \ldots, T_n$, the NDCG is computed as where r(i) is the relevance level of the i[th] tag and Z_n is a normalization constant that is chosen so that the optimal ranking's NDCG score is 1. After computing the NDCG measures of each image's tag list, we can average them to obtain an overall performance evaluation of the tag ranking method.

$$NDCG@k = \frac{1}{z}\sum_{i=1}^{k}\frac{2^{rel(i)} - 1}{\log(1+i)} \quad (7)$$

rel(i) is a binary indicator, which is equal to one if the i[th] tag in the ranking list is relevant to an input image, and zero otherwise.

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1 - Score = \frac{2 * precision * recall}{precision + recall} \quad (10)$$

5. Dataset

Two datasets are used: self-generated and NUS-WIDE. The self-generated dataset consists of images collected from Flickr image sharing website belonging different categories such as fish,

actor, aeroplane, butterfly, autumn etc. and each category consists of 300 images. The size of the images is fixed with maximum width or height to 320 pixels. The eight categories of the images from NUS-WIDE dataset are used for experimentation. The images are divided into two groups: training and testing images after 10-fold cross validation.

For self-generated dataset, the tags associated with the images are collected from Flickr using public API. However, some tags are do not describe the image content. Therefore, tags related to year, camera and brands are excluded from the tag list.

6. Experimental Results

The performance of traditional K nearest neighbor and nearest neighbor is shown in Table 1 and 2 respectively.

Table 1: Performance of traditional K nearest neighbor

Method	Self-generated dataset	NUS-WIDE dataset
Precision	84.33 %	65.87 %
Recall	84.50%	60.75 %
F1-Score	84.00%	61.37 %

Table 2: Performance of Nearest Neighbor

Method	Self-generated dataset	NUS-WIDE dataset
Precision	87.73 %	69.88 %
Recall	88.66 %	67.00 %
F1-Score	88.00 %	67.13 %

From Table 1 and Table 2 it is observed that the performance of nearest neighbor based on each category achieves good performance as compared to traditional k nearest neighbor.

The Figure 4 shows that the NDGC value is high for tag recommendation result when no of the tags recommended is 5 for self-generated dataset and 15 for NUS-WIDE dataset.

Table 3 shows the performance of the existing and proposed tag recommendation algorithm. The performance of proposed tag recommendation algorithm is better as it can recommend/suggest tags with higher NDGC score.

Tagvoting method [18]: In this method, the feature based similar images are determined and the tags are recommended to an input image based on the frequency of tags that appeared in k visually similar images. The method assigns a uniform weight to each neighbor.

TagProp method [20]: In TagProp method, the weights are assigned each visual neighbor of a query image. The weights are assigned using rank based and distance-based method.

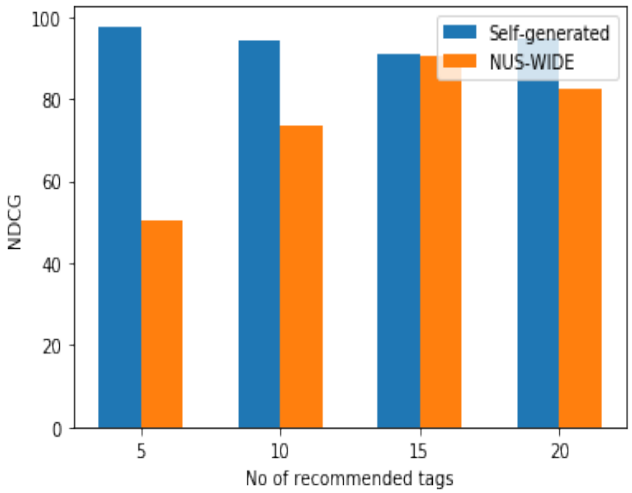

Figure 4: The NDCG values for different number of tags

NVote method [19]: In Nvote method, the tags are recommended based on the difference between global and local tag frequency by assigning equal weight to each neighbor.

Table 3: Performance of different tag recommendation algorithm

Method	Self-generated dataset	NUS-WIDE dataset
Tagvoting	84.45 %	85.20 %
TagProp	89.67 %	68.50 %
NVote	92.07 %	68.96 %
ITR-WTF	95.43 %	88.90%

From Table 3, it is observed that the methods for tag recommendation using the nearest neighbor depend on the value of k which indicates the count of neighbors. The existing methods first identify the uniform/weighted neighbors and then consider only tag information for tag voting which affects the accuracy of tag recommendation. The proposed method improves the accuracy of tag recommendation by combining the image score and tag score.

Table 4: Result of Tag Recommendation

Image	Initial Tags	Recommended Tags
	Brainedge Background Celebration Closeup Clover	Clover Green Nature Leaves Macro

	Bracom (Bram) Bracom Forest Bos Autumn Herfst	Autumn Trees Leaves Path Forest
	Mikepaws London Airport Aircraft Aeroplane	Flying Aircraft Aeroplane Plane Air
	Demerarah Fishalive Tropical-Fish Fish Tropical Fish-Tank Underwater	Fish Aquarium Tropical-Fish Underwater Fishalive

Table 4 shows the result of tag recommendation obtained using the proposed algorithm on self-generated dataset. The initial tag does not describe the entire image content. The proposed algorithm recommends relevant tags to the images. For the second image in table 4 initial tags do not cover the trees, forest and path between the trees which are added by the proposed algorithm.

Therefore, the effectiveness of the proposed method ITR-WTF for tag recommendation is demonstrated using examples in Table 4.

7. Conclusion and future work

In the paper, a method is proposed for tag recommendation of the images by identifying rank neighbors from each category. The method improves the accuracy of tag recommendation by combining the tag frequency score and weighted similarity score of the nearest neighbor images of each category. The experimentation is done two datasets: self-generated and NUS-WIDE dataset. The effectiveness of the proposed method is demonstrated in the experimental results.

The future work will focus on: i) exploring the relationship between tags obtained using from nearest neighbor ii) developing a more optimized approach which works on large dataset iii) exploring metadata associated with the images.

References

[1] E. Spyrou, P. Mylonas, "An overview of flickr challenges and research opportunities," in Proceedings - 9th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP-2014, 2014, doi:10.1109/SMAP.2014.19.

[2] Z. Hao, H. Ge, L. Wang, "Visual attention mechanism and support vector machine based automatic image annotation," PLoS ONE, 2018, doi:10.1371/journal.pone.0206971.

[3] Wu, Wei, Jianyun Nie, and Guanglai Gao, "An improved SVM-based multiple features fusion method for image annotation," Journal of Information & Computational Science **11**(14), 4987-4997, 2014.

[4] X. Li, C.G.M. Snoek, "Classifying tag relevance with relevant positive and negative examples," in MM 2013 - Proceedings of the 2013 ACM Multimedia Conference, 2013, doi:10.1145/2502081.2502129.

[5] H.M. Chen, M.H. Chang, P.C. Chang, M.C. Tien, W.H. Hsu, J.L. Wu, "Sheepdog - group and tag recommendation for flickr photos by automatic search-based learning," in MM'08 - Proceedings of the 2008 ACM International Conference on Multimedia, with co-located Symposium and Workshops, 2008, doi:10.1145/1459359.1459473.

[6] T. Uricchio, L. Ballan, L. Seidenari, A. Del Bimbo, "Automatic image annotation via label transfer in the semantic space," Pattern Recognition, 2017, doi:10.1016/j.patcog.2017.05.019.

[7] X. Qian, X.S. Hua, Y.Y. Tang, T. Mei, "Social image tagging with diverse semantics," IEEE Transactions on Cybernetics, 2014, doi:10.1109/TCYB.2014.2309593.

[8] J. Liu, Z. Li, J. Tang, Y. Jiang, H. Lu, "Personalized geo-specific tag recommendation for photos on social websites," IEEE Transactions on Multimedia, 2014, doi:10.1109/TMM.2014.2302732.

[9] Y. Gao, M. Wang, Z.J. Zha, J. Shen, X. Li, X. Wu, "Visual-textual joint relevance learning for tag-based social image search," IEEE Transactions on Image Processing, 2013, doi:10.1109/TIP.2012.2202676.

[10] L. Zheng, Z. Tianlong, H. Huijian, Z. Caiming, "Personalized tag recommendation based on convolution feature and weighted random walk," International Journal of Computational Intelligence Systems, 2020, doi:10.2991/ijcis.d.200114.001..

[11] Y. Ma, Y. Liu, Q. Xie, L. Li, "CNN-feature based automatic image annotation method," Multimedia Tools and Applications, 2019, doi:10.1007/s11042-018-6038-x.

[12] Q. Ji, L. Zhang, X. Shu, J. Tang, "Image annotation refinement via 2P-KNN based group sparse reconstruction," Multimedia Tools and Applications, 2019, doi:10.1007/s11042-018-5925-5.

[13] V. Maihami, F. Yaghmaee, "Automatic image annotation using community detection in neighbor images," Physica A: Statistical Mechanics and Its Applications, 2018, doi:10.1016/j.physa.2018.05.028.

[14] J. Zhang, Y. Yang, Q. Tian, L. Zhuo, X. Liu, "Personalized Social Image Recommendation Method Based on User-Image-Tag Model," IEEE Transactions on Multimedia, 2017, doi:10.1109/TMM.2017.2701641.

[15] Q. Ji, L. Zhang, Z. Li, "KNN-based image annotation by collectively mining visual and semantic similarities," KSII Transactions on Internet and Information Systems, 2017, doi:10.3837/tiis.2017.09.016.

[16] Y. Verma, C. V. Jawahar, "Image Annotation by Propagating Labels from Semantic Neighbourhoods," International Journal of Computer Vision, 2017, doi:10.1007/s11263-016-0927-0.

[17] S. Lee, W. De Neve, Y.M. Ro, "Visually weighted neighbor voting for image tag relevance learning," Multimedia Tools and Applications, 2014, doi:10.1007/s11042-013-1439-3.

[18] C. Chaoran, Jialie Shen, Jun Ma, and Tao Lian, "Social tag relevance learning via ranking-oriented neighbor voting," Multimedia Tools and Applications **76**(6) 8831-8857, 2017.

[19] X. Qian, X. Liu, C. Zheng, Y. Du, X. Hou, "Tagging photos using users' vocabularies," Neurocomputing, 2013, doi:10.1016/j.neucom.2012.12.021.

[20] L. Xirong, Cees GM Snoek, and Marcel Worring, "Learning social tag relevance by neighbor voting," IEEE Transactions on Multimedia **11**(7), 1310-1322, 2009.

[21] M. Guillaumin, T. Mensink, J. Verbeek, C. Schmid, "TagProp: Discriminative metric learning in nearest neighbor models for image auto-annotation," in Proceedings of the IEEE International Conference on Computer Vision, 2009, doi:10.1109/ICCV.2009.5459266.

[22] D.A. Dondekar and Balwant A. Sonkamble, "Analysis of Flickr Images using Feature Extraction Techniques," 4th IEEE International Conference on Computer and Communication Systems (ICCCS 2019), Singapore, 278-282 2019.

[23] D.A. Dondekar and Balwant A. Sonkamble,"Tag-based Image Retrieval using Hybrid Visual-Tag Feature Extraction Method", International Journal of Advanced Science and Technology, **9**(4), 5931 – 5940, 2020.

[24] X. Li, "Tag relevance fusion for social image retrieval," Multimedia Systems, 2017, doi:10.1007/s00530-014-0430-9.

[25] D.A. Dondekar and Balwant A. Sonkamble, "Harmonic Mean based Classification of Images using Weighted Nearest Neighbor for Tagging" International Journal of Advanced Computer Science and Applications, **11**(11), 2020, doi:/10.14569/IJACSA.2020. 0111131