Tag Recommendation Techniques for Images: A Survey

Anupama D. Dondekar and Balwant A. Sonkamble
Dept. of Computer Engg., PICT, Pune, India
Email: {agphakatkar, basonkamble}@pict.edu

***Abstract*—The users from all localities of the world have given more opportunity to share information with each other. The diverse accessibility of information through various services is useful for developing technologies. These technologies enable the user to share the various variant format of information. In the information sharing, images are observed to be more informative due to its content presence of visualization. The Flickr is an image sharing social website. It allows users to share an image with others for social interaction. The users of the Flickr can assign tags to the images. These tags are used for indexing, organization, and searching of the images. Since tags are assigned by the users, there is a semantic gap between the tags and the actual content of the images. A misclassification or non-relevancy is observed in retrieval when these tags are not properly assigned. This may affect the performance of the image retrieval system and poses new challenges to the image retrieval research community. In this article, we propose techniques for image tag recommendation.**

***Index Terms*—social images, tag recommendation, tagging, annotation, tag ranking, tag relevance**

I. Introduction

Many industries in the world are concentrating on automating the system, which leads to the growth of digital technologies. There is a need for sharing the information through the web and social sites. The information on the website is always in the form of text, images, audio and video. Many image sharing social sites such as Flickr, Panoramio allows users to share images with other users. These websites contain a large volume of image information, which has posed a great challenge for large-scale image search. In image retrieval system text based and content based approach is used to retrieve an image from a large volume of image database. In text [1] based image retrieval approach, the images are retrieved using text query and texts around the images. In content based image retrieval approach, visual similarity is identified between input image and images stored in the database. Both the global and local features of images are used to determine the visual similarity. However, due to semantic gap, the retrieval performance of the content based image retrieval system is limited. To overcome these issues, social tagging was proposed by the researcher with the variety of media based social networking sites such as Flickr, Panoramio.

The Flickr is a well-known website owned by Yahoo! [1]. It is used as a virtual community for image based information sharing and is less motivated with commercial gains. This application has the vast collection of community tagged images and offers a public API to access this images and metadata information. It also enables a user to interact with other users by sharing comments, making favorite about photos and create groups of particular interests, follow other users. More than 14 billions of images have been currently uploaded to Flickr. At the time of uploading images, the users are assigning tags to a particular image as per their view. The tags of the images are specified with visual content or location, or free textual description of an image. These tags can be used to index, classification and retrieval of images. For tag-based image retrieval system, an automatic image tagging plays a critical role.

Existing tagging methods limit the performance of a tagging because of several reasons. The tags assigned by many users are usually limited in free tagging manner. It is due to a difficulty in describing the images with a number of words in a short moment. Also, there is too much noise in user-provided tags. Many tags are irrelevant or incorrectly spelled. As reported in [2], only about 50% of the tags provided by Flickr users are really related to the images. Due to this, the user generated tags rapidly became inconsistent and ambiguous which make the accuracy of image retrievals worse. The retrieval accuracy of such retrieval system is highly dependent on the correlation between user tag and the allocated tag. A misclassification or non-relevancy is observed in retrieval when these tags are not properly assigned. It is hence required to have accurate tagging to an image to have better retrieval accuracy. Toward this requirement, research efforts are made for image tagging, where an automatic tag generation is made for advising appropriate tags to the user. Tag recommendation improves accuracy by suggesting the set of relevant tags for a given image. It can reduce the manual cost of tagging, as it will be faster for users to click on suggested candidates than typing the whole words. Tag recommendation is able to remind the user of alternative tags and also help to clarify the semantics of multimedia content. In addition, it also helps to reduce misspellings and meaningless words.

Currently, the Flickr provides two ranking options for tag based image retrieval. One is time-based and other is interesting-based ranking [3]. In time-based ranking,

Manuscript received August 11, 2017; revised November 17, 2017.

images are ranked based on the uploading time of each image, and the interestingness-based ranking images are ranked based on each image's interestingness in Flickr. These methods do not take the content or visual features and tags of images into consideration. Therefore, both ranking strategies are not based on relevance measure, and thus the search results are not sufficiently good in terms of relevance. Therefore, the retrieval performance of image retrieval can be enhanced by suggesting relevant tags to the user. The Fig. 1 shows typical tag recommendation system for images.

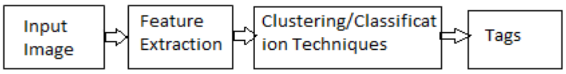

Figure 1. Tag recommendation system for images.

The feature extraction technique is the first step required for an image tag recommendation. Many feature extraction techniques proposed by the researcher are based on color, shape, and texture. After extracting the features of the images, the tags can be assigned using data mining techniques, statistical or probabilistic models. Tags for the can be recommended based on tags, image content, and metadata.

II. Tag Recommendation Techniques

In Flickr, images are associated with the tags, geolocation, contextual information, textual metadata such as title, description, social information. Thus, the techniques for tag recommendation are classified using two criteria: the first criteria uses the information associated with the images and the second criteria determines the methods which are used to exploit the information associated with the images. Based on first criteria, the previous tag recommendation techniques have exploited: i) Previously assigned tags to the images ii) Textual feature (title, description, comments given for the images) iii) Social feature (mark the images of other user as favorites, special interests group, friends) iv) Context feature such as geo location, date, time v) Visual features of the images. Using second criteria, the tag recommendation techniques are classified into following groups: i) Tag co-occurrence based technique which explore the tags assigned previously to the images ii) Hybrid based technique which suggest tags using visual features and tags iii) Graph based technique which model tag recommendation system as graph iv) Matrix factorization based technique which model tag recommendation as matrix v) Clustering based technique which groups different objects (images, user, tags) using different clustering algorithms.

We note that the tag recommendation techniques based on above two criteria are not disjoint. Many tag recommendation techniques uses multiple methods and information associated with the images e.g. some techniques uses visual features, tag features and combine them by means of hybrid technique.

A. Tag Recommendation Based on Tag Co-occurrence

In this technique, for a given image tags are suggested by exploring tag information associated with the images. The tag co-occurrence is calculated using TF-IDF, symmetric and asymmetric measures.

In [2], Borkur Sigurbjornsson *et al.* suggest a tag based on tag repository of all users reside in Flickr. Given a photo with tags, m number of ordered candidate tags are derived for each of the user-defined tags, based on tag co-occurrence. These tags are used as input for aggregation and ranking, which produces n number of ranked tags to be recommended. The recommended tags are used as an index for an image retrieval system. The experiment is conducted on 331 images downloaded from Flickr. The proposed method is good for recommending locations, objects tags. But it is computationally expensive to work on all images and tags. Nikhil Garg *et al.* proposed work to suggest a tag by exploring user's previous tagging history [4]. When a user assigns tags for an image, the system suggests tags based on the tags that have used in user past along with the tags already entered. The user may select tags from suggested list or simply ignore it and add an own tag to an image. The Hybrid algorithm combining conservative performance, classical classification (as Naive Bayes local, tf·idf global, and The Better of two worlds) and new cost measure is used in the proposed work. The advantage of proposed work is the low average cost of the tagging. The work did not explain about the goodness of recommended tag. Aixin Sun *et al.* proposed the knowledge-based approach for image tag recommendation which exploits tag concepts [5]. Tag concepts are derived based on tags of images represented in the form of tag co-occurrence pairs. The tag relationship graph is constructed for each candidate tag. Then modularity clustering is used to identify the concept associated with each tag. These concepts are used as an index. The matching concepts are retrieved from the index using cosine similarity. Finally, the candidate tags associated with a matching candidate is recommended. The method boosts scalability and efficiency of tag recommendation process. However, it does not consider features of the images to improve visual search. Luca Cagliero *et al.* [6] developed a method to suggest tags based on generalized association rules which explore tags correlation holding at different abstraction levels to identify additional tags for a suggestion. The method store tag set in a transactional data and a generalization hierarchy G_H is built over transaction tag. Given the minimum threshold and support, GENIO algorithm is used over G_H to generate user specific rule set and collective rule set. Using these rules, tags are selected and are ranked using Borda Count Consensus function. Due to the generalized rule set, the performance of the recommender system is very high. But the proposed system does not work well for the images with noisy tags. Rae *et al.* proposed tag recommendation system to suggest the additional tag for partially annotated images [7]. The recommended system recommend tags using utilizing the knowledge that present at four different contextual layers using the probabilistic framework. The first contextual layer is personal context, constructed from the annotations provided by the user. Secondly, a social contact context is defined by aggregating the annotations over all users that are identified as a contact of that user. Thirdly, a social group context is obtained by aggregating the photo annotations of photos posted in the groups that the user is subscribed to. Finally, a collective context is determined by aggregating the annotations for all photos posted by all users. A network of tags is derived for each context, based on co-occurrence analysis of tags used to annotate the photos within that context. To produce a set of recommendations for a given set of input query tags, each query tag is used to generate an intermediate set of recommendations and these sets are then combined. For each of the contexts, the same probabilistic model and Borda Count based aggregation approach are used to generate recommendations from different contexts into a unified ranking of recommended tags. The method is able to recommend relevant tags to an image irrespective of language. Xian Chen and Hyoseop Shin proposed an algorithm to suggest tag using both textual and social features [8]. Through both textual analysis and social activities, the algorithms collect features for tags of each user. To find tags which are related to user's interest textual features are extracted from each user's own terms in tags, titles, contents, and comments which are applied to his/her own photos. Also, the user favorite topic is determined from some tags used frequently. Social features are determined from each user's social activities e.g. a user marks other users' photos as his/her favorites, or another user marks his/her own photos as favorites. If users marked other several photos which are related to the same tags, the algorithm infers that users are interested in the topics on these tags. Naive Bayes classifier is used to find the representative tags that can be related to the users' favorite topics. The algorithm is good at finding high-quality tags for items.

The advantage of the technique is: it is easy to compute and previously assigned tags are easily available. But these tags are highly personalized and may not be relevant to images. The solution is to use image features along with tags for accurate tag recommendation.

B. Hybrid Tag Recommendation

For given image, the technique uses both visual information and associated tags. It is based on the idea that visually similar images are labeled with similar tags. The technique can be implemented in three different ways. First, is to explore images which are visually similar to the query image. Second, is to determine the similarity between images labeled with same tags. Third, is to build the classifier using socially tagged examples. These three methods find an image similarity and tag similarity sequentially or by finding a common structure to link tag and image information.

In [9], S. Lee *et al.* presented image tag recommendation as a maximum posteriori problem by making use of a visual folksonomy. The visual folksonomy contains information of user, tags, images and ternary relationship between users, tags and images. Given the folksonomy F, the algorithm first constructs a new folksonomy Q that contains visually similar images to the input image q using MPEG-7 color and texture descriptor. Second, to reduce the number of images that are not semantically related to the input image q, select a subset of images in Q that are all related to a particular tag t based on probability . For visually similar images in R, the algorithm recommends top 10 meaningful tags for the query image q using a posteriori probability. Tag to tag relationship is not considered in the proposed method. In [10], H. Chen *et al.* implemented SVM classifier to recommend tags and group for images. The method first extracts the color, texture and edge features of the images and trained SVM classifier to detect 62 concepts. The SVM predictor gives each concept a probability value to indicate the degree that an image I fits this concept. The probabilities of each concept are then ranked in the descending order. The top n concepts are extracted from ranked result. For each of the top n concepts predicted for an image I, the method recommend top p suitable groups, and discover popular tags by gathering the statistics of the tags which is attached to any image in these groups. In the proposed work, the data gathering consumes more time and adding new concept takes more time for reconstructing the classifier. In [11], Jing Liu *et al.* proposed method for personalized tag suggestion by jointly exploring the tagging resources and the geo-location information in social web context. The user preferences is determined from tagging history and analyze the geo-location preference based on the location related tagging resources using unified subspace learning method. A unified subspace is shared by the visual and textual features. The visual feature is a lower level representation of semantics than the textual information, which adopts a progressive learning strategy for consistent representation. For given an untagged photo of a user, the method map untagged photo into the unified space corresponding to the user and use the nearest neighbour search to obtain some user-preferred tags and geo-location-preference tags individually. The tags and the visual features of the photo are combined to identify semantically and visually related photos. Finally, the top-ranked tags are recommended to the user. By restricting the images to be taken within the same city, the subspace carries geo information to some extent. However, the city level resolution is too coarse to precisely describe the geo context of a specific image. X. Li *et al.* proposed method which suggests the tags for labeled as well as unlabeled images [12], [13]. The relevance of a tag with respect to an image is determined from tagging behaviour of visual of that image. The method identifies the similar images based on visual features solely using KNN classifier. It does not impose any model training for any visual concepts. Thus, it is scalable to handle a large amount of social-tagged images. But, it is not able to explore the correlation between tags to estimate visual similarity between images. It uses only one single feature to determine the neighbor images based on visual features. The accuracy can be improved using multiple features. In [14], to solve the problem of learning tag relevance for multiple features, two simple and effective methods based on classical Borda Count and Uniform Tagger is suggested. Both methods combine the output of many tag relevance learners driven by diverse features in an unsupervised manner, rather to supervised, manner. The method is used for tag ranking, tag enrichment, and tag-based image retrieval. Experiment is carried out on two datasets: NUS-WIDE and dataset created by author. The methods show 74% retrieval accuracy using improved tags. But it need more time for large dataset and the dataset with noisy labels. Xueming Qian *et al.* presented an approach to retag social images with diverse semantics [15]. The relevance of a tag to an image is determined using three approaches. The first approach used is KNN classifier to map the tag from textual space to low-level feature space by representing the tag with a set of images containing the tag. Then, the visual relevance of a tag to image can be measured by the image-image similarity in low-level visual feature space. In the second approach the semantic relevance of a tag to the image is measured by Google distance formula. In the third approach both the visual relevance in low-level visual spaces and the semantic relevance in high-level textual spaces are combined. The diversity of the tag to the image is measured by the product of the relevance of the tag to image and its semantic compensation to the tags ranked ahead of it. Finally, an iterative greedy searching based tagging approach is used to determine the optimal tag list. The method shows improvement in textual based image retrieval, filtering noise samples from a large scale weakly labeled image set. But it does not define any approach to tag variation for variations in image content. Cui *et al*. proposed method to suggest tags for an image based on visual features and tag correlations using neighbor voting scheme [16]. The color, texture features are used to find visual neighbors of a given image. All features are concatenated into a single vector and Euclidean distance is used to find the similarity between images. The pairwise relationship between tags is determined as low-rank approximation problem. The limitation of the algorithm is, it require more time for training SVM classifier. Xueming Qian *et al.* [17], [18] proposed method based on user's history information. Let M denote the number of the total uploaded images by the user u. The user u history contains visual features, time, and geographical coordinates of the images uploaded by the same user. For a given image, the algorithm finds the images using three parameters: visually similar images, images with same GPS coordinates and images which are taken with a period of time from the user's history. The tags discovered in these three neighbors are collected together and count their frequency by voting. These frequently appeared tags were used to recommend the newly uploaded image. The algorithm is not able to recommend tag if the there are no candidate image of the input image and highly dependent on geographical coordinates and times they were taken. Also, the algorithm used visual, geo, and temporal for personalized image tagging, without considering tag feature. In [19], Emily Moxley *et al.* suggest tag re-ranking based on geolocation and content based image analysis. When a user adds a new image along with an accompanying latitude-longitude coordinate pair, SpiritTagger extract images within a certain geographic area from the target image using N nearest neighbor algorithm. A candidate set of tags is then collected, and calculate the global and local tag distribution for geographically representative tags seen frequently in similar images. The highest scoring tags can then be suggested to the user for annotating his/her upload. The tags are recommended for landmark images of Los Angels and Southern California region. The algorithm does not work properly for a larger region. Rawat *et al.* proposed deep neural network to suggest multiple tags for a given image [20]. The model is trained using visual features and context information such as time, geolocation of the images. The method shows 83% accuracy in the tag recommendation. The experiment is conducted on YFCC100M image dataset. In the proposed work the correlation between multiple tags based on user context is not considered for tag recommendation.

The main advantage of the technique is the accuracy of tag recommendation is improved by combining the visual content of the images and tags. But, the technique is not able to handle image diversity problem, images which are different in content but visually appears the same.

C. Graph Based Tag Recommendation

In this technique, the node of the graph corresponds to image, tag, user or group information and there is an edge between nodes if they are similar. The similarity is calculated in terms of textual similarity between tags or visual similarity between images or both.

In [3], Liu *et al.* proposed tag ranking method in which relevance score for individual tags associated with an image are assigned based on Kernel Density Estimation (KDE) initialization and random walk refinement. For given an image along with its tags, a probabilistic approach is used to calculate initial relevance score of each tag. Tag relevance determined using probabilistic score does not consider the tag relationship. The relationship between tags is determined using random walk-based refinement to increase the performance of tag ranking. Finally, the ranking of the tags of the images is done according to their refined relevance scores. In the proposed method, the image similarity based only on visual features is unable to consider tags into account and does not able to capture semantic relationship among the images. Jing Liu *et al.* proposed graph learning framework for image annotation [21], [22]. The framework contains two graph models: the image based graph model and the word based graph model. To annotate an image, image based similarity graph is constructed using nearest spanning chain method to propagate labels from the annotated images to the un-annotated images by their visual similarities. The visual similarity is determined using color and texture features. The word-based graph learning was performed to refine the relationships between images and words by exploring three kinds of word correlations. One is the word co-occurrence in the training set, and the other two are derived from the web context. The technique works on small label dataset, but their performance degrades when the dataset size is increased. The Corel data sets and one web image dataset is used to test the algorithms. The problem of annotating a large-scale image, by label propagation over noisily-tagged web images is suggested in [23]. To annotate the images more accurately, a "kNN-sparse graph-based", semi-supervised learning approach for segregation of labeled and unlabeled data simultaneously is suggested. The sparse graph constructed by one-vs-kNN sparse is suggested to remove most of the semantically-unrelated links among the data, and thus it is more robust and discriminative than the conventional graph approaches. In [24], Konstantinos Pliakos *et al*. proposed the method for image tagging and geo-location prediction based on hypergraph learning. It fully exploits the image content, the context, and the social media information. The set of objects is split into different object groups (images, users, social groups, tags, and geo-tags) and each object group effect in image tagging and geo-location prediction is controlled separately by assigning them different weights. In the method, 100 nearest neighbors to each image are identified using GIST descriptor and reduced to 5 most similar images using SIFT. The method achieves 83% accuracy. Xiaoming *et al*. implemented social image tagging as a "ranking and reinforcement" problem and a novel graph based reinforcement algorithm for interrelated multi-type objects are proposed in [25]. When a user u issues a tagging request for a new image I_q the method retrieves a set of visually similar images for image I_q and a set of friends of user u. The visually similar images are clustered using K-means algorithm and are ranked by their semantic consistency value. The semantic consistency value of a retrieved image is estimated based on both of its semantic and visual similarity with other images in the retrieval result. The semantic similarity between each two images I_i and I_j is measured using cosine similarity between the vectors of their tags and terms extracted from surrounding text. Then a candidate tag set T_c is collected from those associated with the neighbour images. A graph G is constructed which consist of three types of objects i.e., visual features extracted from query image I_q, candidate tags, friend users. Each type of object was initially ranked based on their weights and intra-relation. By utilizing inter-relation between the object with different types, the reinforcing process is performed on the feature graph G to re-rank all the objects, and only the top ranked tags of T_c are reserved as final tags. The algorithm provides high accuracy, but determining the relationship between tags and features is not easy. Also, the algorithm does not consider diversity in the images. The two data sets NUS-WIDE and PerFlickr are used for an experiment which consists of contact information collected using Flick API.

The disadvantage of the technique is inability to handle changes in the tagging system. The model retraining is needed when new image or tag is added.

D. Matrix Factorization Based Tag Recommendation

In this technique, the tag assignment problems are model as a matrix and apply dimensionality reduction method on the matrix. Generally, tensors are used to capture the relationship between images, tags and users and the tags are recommended using Tensor factorization method.

Dimitrios Rafailidis *et al*. suggest tag using Tensor Factorization and tag Clustering model [26]. The model consists of three steps. In the first step, relevance feedback mechanism is used to perform tag propagation between similar images based on visual features. In the second step, tag clustering is performed to find topics on the social network using two different tag clustering algorithms: tripartite network and adaptive K-means. The tf·idf weighting scheme is used to identify the interest of users and image relations to topics. In the third step, high order singular value decomposition is exploited to reveal the associations among users, topics, and images. The limitation of the method is, the high order singular value decomposition requires more time for building the recommendation model at the offline part. An automatic way for finding the optimal number of tag clusters is required. In [27], the approach of tag recommendation over a heterogeneous network is developed. The issue of co ranking for tag recommendation was defined as a Bergman divergence optimization, by defining a random walk approach to an equivalent optimal kernel matrix learning problem. The co-ranking objective for tagging over such distributed network is observed to optimize by solving the matrix learning problem. The approach gives an efficient performance for image ranking and tag ranking but reduces the accuracy due to PCA. In [28], Pantraki *et al*. proposed an automatic image annotation and recommendation system based on Parallel Factor Analysis 2. The system was applied to three matrices, namely the image-feature matrix, the visual appearance of images, and the image-tag matrix. To make a tag recommendation, the test image is multiplied by singular vectors of the image-tag matrix, yielding suggesting tag vector. Wang Yong-Sheng proposed algorithm for image tag recommendation using tensor factorization [29]. The image, tag and person matrix are intergraded into one tensor and tags are recommended through tag ranking using tensor factorization method.

The advantage of the technique is dimensionality reduction. But the method is not scalable; it is computationally expensive to build the model.

E. Clustering Based Tag Recommendation

In this technique, to recommend tags clustering algorithms are used to cluster different objects such as images, tags and users and recommend the representative tag of the cluster.

In [30], Haifeng *et al*. proposed method to suggest image tag recommendation using visual features of images and friendship information among users in the social network. In the first step, the similarity between the images is calculated based on visual features. In the second step, a biclustering algorithm is used calculate the pair-wise user similarities and the similarities among tags. In the third step, a maximum a posteriori is used to suggest the tags for the query image. The proposed approach suggests tags for images from MIRFlickr and

PISAR dataset. The tag-tag relationship is not considered to recommend the high quality tags. In [31], web browsing behaviour of the user is exploited to suggest tags for images which not only to add but also to be deleted for user. The web pages are clustered using hierarchical clustering algorithm and tag frequency is calculated using TF-IDF within each cluster. The tag appears frequently within each cluster is called as representative of the cluster and the relevance between two tags is calculated using Normalized Google Distance.

The technique reduces the dimensionality of the problem, but the candidates may be too general to describe the specific content of the target image or discriminate it from others.

III. Conclusion

Tag recommendation is very important and challenging task, which helps users to assign appropriate keywords to an image, which in turn help to retrieve images from a huge collection of image database. Many researchers have created own data set and done extensive research work to develop new algorithms for tag recommendation. Still, there are some issues which are open. First, the quality of the metadata is low; many tags are ambiguous and personalized. Second, in the social network images are added at random interval, the web scaling problem needs to consider in tag recommendation. Third, due to diversity in the images, it is necessary to develop an approach to reduce false tagging issue. Finally, improve the tagging accuracy and performance by developing a new approach of fusion technique of image features and tag and user information.

References

[1] E. Spyrou and P. Mylonas, "An overview of flick challenges and research opportunities," in *Proc. IEEE International Workshop on Semantic and Social Media Adaptation and Personalization*, 2014, pp. 88-93.

[2] B. Sigurbjonsson and R. V. Zwol , "Flickr tag recommendation based on collective knowledge," in *Proc. ACM International Conference on World Wide Web*, 2008, pp. 327-336.

[3] D. Liu, X. Hua, L. Yang, M. Wang, and H. Zhang, "Tag ranking," in *Proc. ACM International Conference on World Wide Web*, 2009, pp. 351-360.

[4] N. Garg and I. Weber, "Personalized interactive tag recommendation for Flickr," in *Proc. ACM Conference on Recommender Systems*, 2008, pp. 67-74.

[5] A. Sun, S. S. Bhowmick, and J. A. Chong, "Social image tag recommendation by concept matching," in *Proc. ACM International Conference on Multimedia*, 2011, pp. 1181-1184.

[6] L. Cagliero and P. D. Torino, "Personalized tag recommendation based on generalized rules," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 1, pp. 12:1-12:22, 2013.

[7] A. Rae, B. Sigurbjrnsson, and R. V. Zwol, "Improving tag recommendation using social networks," in *Proc. Adaptivity, Personalization and Fusion of Heterogeneous Information*, 2010, pp. 92-99.

[8] X. Chen and H. Shin, "Tag recommendation by machine learning with textual and social features," *Journal of Intelligent Information Systems*, vol. 40, no. 2, pp. 261-282, 2013.

[9] S. Lee, W. D. Neve, K. N. Plataniotis, and Y. M. Ro, "Map-based image tag recommendation using a visual folksonomy," *Pattern Recognition Letters*, vol. 31, no. 9, pp. 976-982, 2010.

[10] H. Chen, M. Chang, P. Chang, M. Tien, W. Hsu, and J. Wu, "SheepDog: Group and tag recommendation for Flickr photos by automatic search based learning," in *Proc. ACM International Conference on Multimedia*, 2008, pp. 737-740.

[11] J. Liu, Z. Li, J. Tang, Y. Jiang, and H. Lu, "Personalized geo-specific tag recommendation for photos on social websites," *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 588-600, 2014.

[12] X. Li, C. Snoek, and M. Worring, "Learning social tag relevance by neighbor voting," *IEEE Transactions on Multimedia*, vol. 11, no. 7, pp. 1310-1322, 2009.

[13] X. Li, C. Snoek, and M. Worring, "Learning tag relevance by neighbor voting for social image retrieval," in *Proc. ACM International Conference on MIR*, 2008, pp. 180-187.

[14] X. Li, C. Snoek, and M. Worring, "Unsupervised multi-feature tag relevance learning for social image retrieval," in *Proc. ACM International Conference on Image and Video Retrieval*, 2010, pp. 10-17.

[15] X. Qian, X. S. Hua, Y. Y. Tang, and T. Mei, "Social image tagging with diverse semantics," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2493-2508, 2014.

[16] C. Cui, J. Shen, J. Ma, and T. Lian, "Social tag relevance learning via ranking-oriented neighbor voting," *Multimedia Tools and Applications*, vol. 76, no. 6, pp. 1-27, 2017.

[17] X. Qian, X. Liu, C. Zheng, Y. Du, and X. Hou, "Tagging photos using users' vocabularies," *Neurocomputing*, vol. 111, pp. 144–153, 2013.

[18] X. Liu, X. Qian, D. Lu, X. Hou, and L. Wang, "Personalized tag recommendation for flickr users," in *Proc. IEEE International Conference on Recommender System*, 2014, pp. 1-6.

[19] E. Moxley, J. Kleban, and B. S. Manjunath, "Spirittagger: A geo-aware tag suggestion tool mined from Flickr," in *Proc. ACM International Conference on Multimedia Information Retrieval*, 2008, pp. 24-30.

[20] Rawat, Y. Singh and M. S. Kankanhalli, "ConTagNet: Exploiting user context for image tag recommendation," in *Proc. ACM International Conference on Multimedia Conference*, Oct 2016, pp. 1102-1106

[21] J. Liu, M. Li, W. Ma, Q. Liu, and H. Lu, "An adaptive graph model for automatic image annotation," in *Proc. ACM International Workshop on Multimedia Information Retrieval*, 2006, pp. 61-70.

[22] 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.

[23] J. Tang, R. Hong, *et al.*, "Image annotation by kNN-sparse graph-based label propagation over noisily-tagged web images," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 2, pp. 11-125, 2010.

[24] K. Pliakos and C. Kotropoulos, "Simultaneous image tagging and geo-location prediction within hypergraph ranking framework," in *Proc. IEEE International Conference on Acoustic, Speech and Signal Processing*, 2014, pp. 6894-6898.

[25] X. Zhang, X. Zhao, *et al.*, "Social image tagging using graph-based reinforcement on multi-type interrelated objects," *Signal Processing*, vol. 93, no. 8, pp. 2178-2189, 2013.

[26] D. Rafailidis and P. Daras, "The TFC model: Tensor factorization and tag clustering for item recommendation in social tagging systems," *IEEE Transactions on System, Man, Cybernetics: Systems*, vol. 43, no. 3, pp. 673–688, 2013.

[27] L. Wu, X. Huang, C. Zhang, J. Shepherd, and Y. Wang, "An efficient framework of bregman divergence optimization for co-ranking images and tags in a heterogeneous network," *Multimedia Tools and Applications*, vol. 74, no. 15, pp. 5635-5660, 2015.

[28] P. Evangelia, and C. Kotropoulos, "Automatic image tagging and recommendation via PARAFAC2," in *Proc. IEEE Workshop on Machine Learning for Signal Processing*, 2015, pp. 1-6.

[29] Y. Wang, "Image tag recommendation algorithm using tensor factorization," *Journal of Multimedia*, vol. 9, no. 3, pp. 416-422, 2014.

[30] H. Guo, S. Su, and Z. Sun, "Image tag recommendation based on friendships," *Multimedia Tools and Applications*, 2016, pp. 1-17.

[31] T. Taiki, T. Itokawa, T. Kitasuka, and M. Aritsugi, "Tag recommendation for flickr using web browsing behavior," in *Proc. Computational Science and Its Applications*, 2010, pp. 412-421.

Anupama D. Dondekar received B.E. degree from Dr. Babasaheb Ambedkar Marathwada University, Auragngabad and M.E. degree from in Savitiribai Phule Pune University. She is currently a Ph.D. candidate in Savitiribai Phule Pune University. Her research interests include Image Processing, Data Mining and Machine Learning.

Balwant A. Sonkamble currently working as Professor in PICT, affiliated to Savitribai Phule Pune University, Pune, India. He received PhD in Computer Science and Engg. from S.R.T.M.U., Nanded in 2013. His research interests include Artificial Intelligence, Machine Learning, and Speech Processing.