International Journal of MC Square Scientific Research Vol.9, No.1 April 2017

TAG BASED IMAGE SEARCH BY SOCIAL RE-RANKING IN THE WEB BASED APPLICATIONS

VajjaNarendraNath[1]
[1]*UG student,*
Department of Computer Science Engineering,
SRM University, Chennai –89
Mail.id-narendranath9595@gmail.com

SasidharVegi[2]
[2]*UG student,*
Department of Computer Science Engineering,
SRM University, Chennai – 89
Mail.id-sashidharvegi2509@gmail.com

Abstract— Social media sharing websites like Flickr allow users to annotate images with free tags, which significantly contribute to the development of the web image retrieval and organization. Tag-based image search is an important method to find images contributed by social users in such social websites. However, how to make the top ranked result relevant and with diversity is challenging. Proposing a social re-ranking system for tag-based image retrieval with the consideration of image's relevance and diversity is done. Proposed system aims at re-ranking images according to their visual information, semantic information and social clues. The initial results include images contributed by different social users. Usually each user contributes several images. First sorting the images by inter-user re-ranking is done. Users that have higher contribution to the given query rank higher. Then sequentially implementing intra-user re-ranking on the ranked user's image set, and only the most relevant image from each user's image set is selected. These selected images compose the final retrieved results thus building an inverted index structure for the social image dataset to accelerate the searching process. Experimental results on Flickr dataset show that our social re-ranking method is effective and efficient in the real world.

Keywords—Re-ranking, image set**,** data set.

1. Introduction

The existing system is used to learn the relevance of tags by visually weighted neighbour voting, a variant of the popular baseline neighbour voting algorithm. Relevance tag ranking algorithm, which can automatically rank tags according to their relevance with the image content, is being used at present. A modified probabilistic relevance estimation method is proposed by taking the size factor of objects into account and random walk based

refinement is utilized. A tag fusion method is used for tag relevance estimation, to solve the limitations of a single measurement on tag relevance. Besides, early and late fusion schemes for a neighbour voting based tag relevance estimator are conducted. Then an adaptive teleportation random walk model on the voting graph which is constructed based on the images relationship to estimate the tag relevance. Then, a tag clarity score measurement approach to evaluate the correctness of a tag in describing the visual content of its annotated images in the web where Tag mismatches occur. Social tagging requires all the users in the social network to label their uploaded images with their own keywords and share with others. Different from ontology based image annotation; there is no predefined ontology or taxonomy in social image tagging. Every user has own habit to tag images. Even for the same image, tags contributed by different users will be of great difference.

Thus, the same image can be interpreted in several ways with several different tags according to the background behind the image. Thus, many seemingly irrelevant tags are introduced. Query ambiguity. Users cannot precisely describe their request with single words and tag suggestion system always recommend words that are highly correlated to the existing tag set, thus add little information to a users' contribution. Besides, polysemy and synonyms are the other causes of the query ambiguity Then there comes a tag-based image search approach with social re-ranking which systematically fuse the visual information, social user's information and image view times to boost the diversity performance of the search result. Therefore, the inter-user re-ranking method and intra-user re-ranking method are implemented to achieve a good trade-off between the diversity and relevance performance.

These methods not only reserve the relevant images, but also effectively eliminate the similar images from the same user in the ranked results. In the intra-user re-ranking process, fusing the visual, semantic and views information into a regularization framework to learn the relevance score of every image in each user's image set. To speed up the learning speed, Using of the co-occurrence word set of the given query to estimate the semantic relevance matrix is done. In order to improve the robustness of the algorithm to obtain the co-occurrence word set with respect to the given query, a new self-adaptive algorithm is discussed in this paper, in which relative frequency of each tag about the given query is required and a self-adaptive parameter is decided by this relative frequency. In the intra-user, the re-ranking process is considering the views to learn the relevance score of each image in the web. This is done in order to achieve the relevance score. The system is more considerate when compared to existing systems. Discussions about weight selection and image features in the regularization framework are complemented. Through this discussion, the performance doesn't rely on the adjustment of parameters and feature selection.

It's robust and relatively stable. Besides, in order to find an optimal number of representative images which are selected from each user's image set, many new comparison experiments and comprehensive discussions are added.

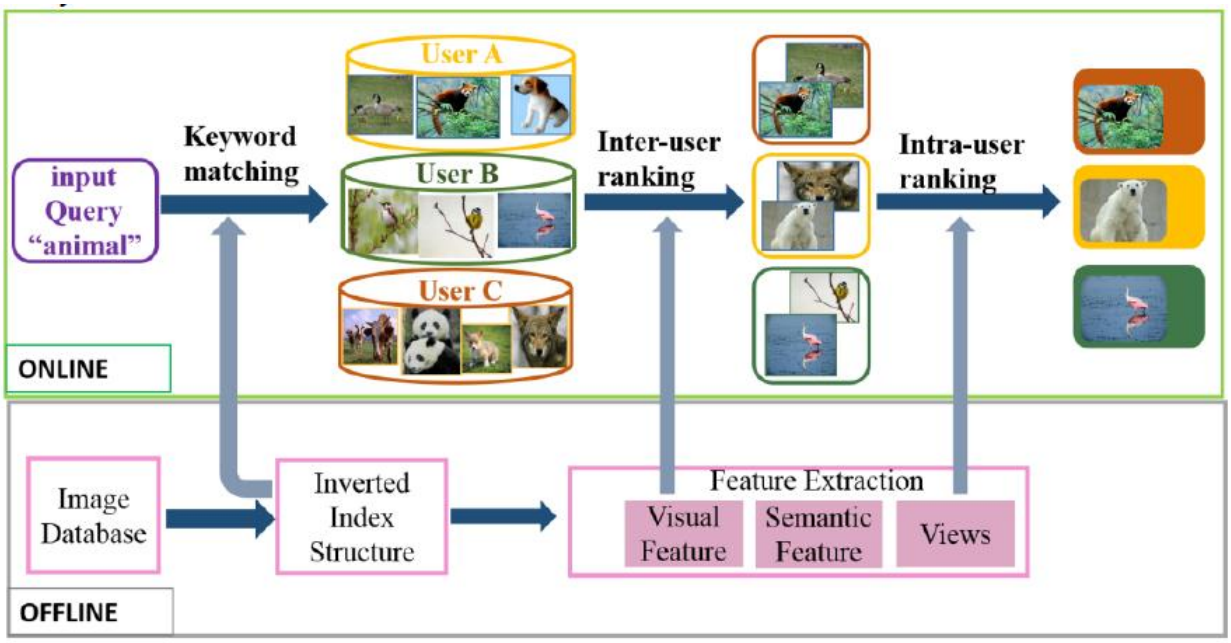

Fig. 1. Basic concept of ranking web pages

2. Literature Survey

Neighbour voting protocol

The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, *k* is a user-defined constant, and an unlabelled vector (a query or test point) is classified by assigning the label which is most frequent among the *k* training samples nearest to that query point. A commonly used distance metric for continuous variables is Euclidean distance. For discrete variables, such as for text classification, another metric can be used, such as the overlap metric (or Hamming distance). In the context of gene expression micro-array data, for example, *k*-NN has also been employed with correlation coefficients such as Pearson and Spear man.[3] Often, the classification accuracy of *k*-NN can be improved significantly if the distance metric is learned with specialized algorithms such as Large Margin Nearest Neighbour or Neighbourhood components analysis. A drawback of the basic "majority voting" classification occurs when the class distribution is skewed. That is, examples of a more frequent class tend to dominate the prediction of the new example, because they tend to be common among the *k* nearest neighbours due to their large number.[4] One way to overcome this problem is to weight the classification, taking into account the distance from the test point to each of its *k* nearest neighbours. The class (or value, in regression problems) of each of the *k* nearest points is

multiplied by a weight proportional to the inverse of the distance from that point to the test point. Another way to overcome skew is by abstraction in data representation. For example, in a self-organizing map (SOM), each node is a representative (a centre) of a cluster of similar points, regardless of their density in the original training data. *K*-NN can then be applied to the SOM.

3. Proposed Method

As illustrated in Fig.2, the system model of re-ranking web pages is given below.

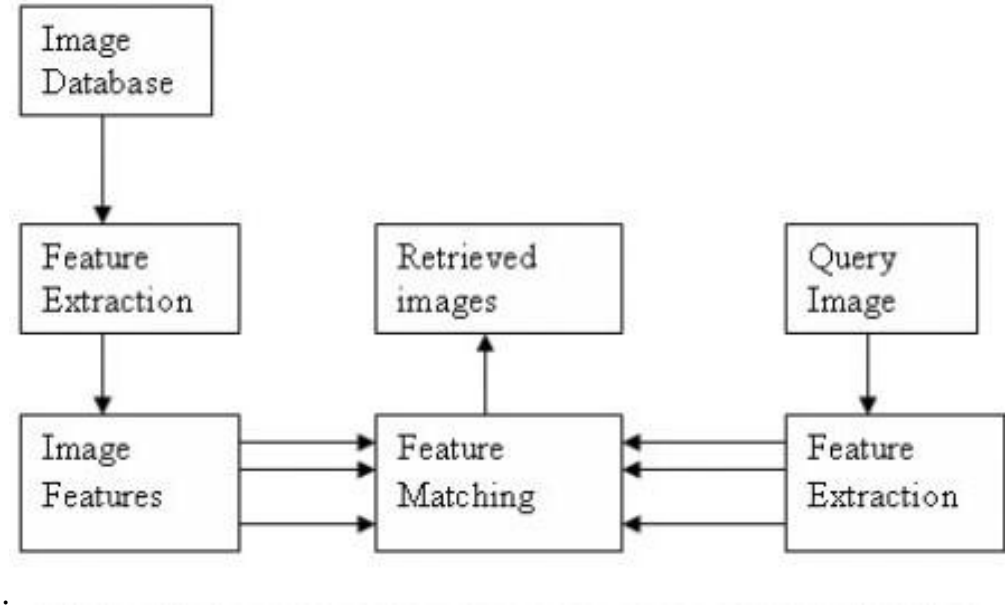

Fig. 2. System model of re-ranking web pages

IMAGE DATABASE

This keeps all the information regarding the dataset information of the clients and the hosts, and this maintains the redundancy of the data in the database.

FEATURE EXTRACTION

This selects the existing feature of the all the available details in the data set uploaded by the user, this can be configured in net beans and selecting the feature that are going to be used further by the others in the web

IMAGE FEATURES

The each and every small details of the picture uploaded in the web are noticed and are stored in the database that is referred when a user comes with a query. Thus by storing this becomes easy for further evaluation of the data sets.

FEATURE MATCHING

This notices all the features of either data sets or the image sets for the uploaded information. The matched features are used when the user comes with a query for the data sets.

FEATURE EXTRACTION

The small details of the data sets are stored in the database and are referred. Whenever, it comes toa use with them by the image query.

QUERY IMAGE

The query image is given by the user who needs to find the relevant searches with the uploaded data sets or the image sets. These checks with all the available image sets of information in the database and the results are derived using the database information.

4. Modules

In this Proposed System there are 4 different modules which are:

- System Construction Module
- Inverted Index Structure Construction
- Feature Extraction
- Obtain the ranked imagesets

System Construction Module

In this module, developing the proposed system module with the entities of User and Admin such that the system extract the visual feature, semantic feature and views for the images data set is done. Semantic feature refers to the co-occurrence word set of query tags and the tags of the images. The new user should initially register with the system and then the user authorization is provided in the User Login. After the authorization, the user can able to upload the images and can set the tags are developed. The module is developed in such a way the user can able to view the Related Images based on Colour and Related Images. It is also designed with the features of user can search the image and can view the image ratings based on Highest Ratings and Lowest Ratings. In the admin entity, is developed as such the admin can able to view the User Registration details and can upload images. Admin is provided with

the facility of view all the users, uploaded images and check the Words, Tag and Duplication etc. If the admin rejects the image, then it can be viewed.

Inverted Index Structure Construction

To realize fast retrieval, an inverted index structure for the collected images is built. In this experiment, our image dataset is composed of the images uploaded by all the users. Each user has uploaded several images. The organization form of original images is based on users. And the inverted index structure is based on tags and each tag corresponds to the images uploaded by different users.

Feature Extraction

In this module use of the visual features, views and semantic features to represent the images in image dataset is done. Color feature is one of the most widely used visual features in image retrieval, for its in variance with respect to image scaling, rotation, translation and also an image is divided into four equal sized blocks and a centralized image with equal-size. The 9-D color moment of an image segment is utilized, which contains values of mean, standard deviation and skewness of each channel in HSV color space. Texture feature describes the structure arrangement of surfaces and their relationship to the environment, such as fruit skin, clouds, trees, and fabric. The texture feature in the proposed method is described by hierarchical wavelet packet descriptor (HWVP). A HWVP descriptor is utilized by setting the decomposition level to be 3 and the wavelet packet basis to be DB2.

Obtain the ranked images

The views of an image in social media community are an important feature which indicates the click count of this image. The number of click count has been utilized to improve the relevance performance of the image retrieval results. Besides, clicks have also been used to estimate the documents relevance. Co-occurrence is a linguistics term that can either mean concurrence/coincidence. In a more specific sense, co-occurrence means two terms which often appeared in the text corpus in a certain order. It can also be interpreted as an indicator of interdependency, semantic proximity or an idiomatic expression and often be used in the study of image tagging. The proposed online system carries out the following three steps to obtain the ranked images for the query tags: 1) keyword matching, 2) inter-user re-ranking, and 3) intra-user re-ranking. In the keyword matching the keywords from the user input and the also from the web will be checked whether they are matched or not. In inter user re-ranking the keywords or the pictures from the different user of the web or a network are identified and re-ranking is done on that basis. In intra user re-ranking the content of the same user that was put on the web is re-ranked based on the priorities.

4.1 Implementation of Modules

ALGORITHM

Input: The image sets or data sets.
Output: Re-ranked pages of the users.

1. USER A uploads and tags some image or datasets.
2. USER C….Z also tags the relevant or irrelevant data sets for USER an uploaded datasets.
3. USER B searches the web which slightly relates to User A
4. Due to the tagging given by USER C…..Z, User B obtains the user an uploaded datasets

5. Conclusion& Futurework

By the simulation results the conclusion can be drawn that the proposed system can be implemented in the real world in order to analyze the re-ranking of the web pages. The proposed system can be used give the other user rights to tag the features that will be used to notify even the small features in the datasets in the web pages. The system can be implemented on the simulation platform which can be net beans and creating and building database files for the image sets and the data sets. The database files are first noted with the featured images in the database in order to notify the query results for the users searching the data sets in the web applications.

References

[1] D. Liu, X. Hua, L. Yang, M. Wang, H. Zhang, "Tag ranking", Proc. Int. Conf. World Wide Web, pp. 351-360, 2009.

[2] X. Li, C. Snoek, M. Worring, "Learning tag relevance by neighbour voting for social image retrieval", Proc. ACM Int. Conf. Multimedia Inform. Retrieval, pp. 180-187, 2008.

[3] XuemingQian, Member, IEEE, Dan Lu, and Xiaoxiao Liu, “Tag Based Image Search by Social Re-ranking”,IEEE TRANSACTIONS ON MULTIMEDIA, 2016.

[4] D. Liu, X. Hua, M. Wang, H. Zhang, "Boost search relevance for tag-based social image retrieval", Proc. IEEE Int. Conf. Multimedia Expo., pp. 1636-1639, 2009

[5] K. Yang, M. Wang, X. Hua, H. Zhang, "Social image search with diverse relevance ranking", Proc. Int. Conf. Magn. Magn.Mater., pp. 174-184, 2010..

[6] R. Leuken, L. Garcia, X. Olivares, R. Zwol, "Visual diversification of image search results", Proc. 18th Int. Conf. World Wide Web, pp. 341-350, 2009.

[7] R. Cilibrasi, P. Vitanyi, "The Google similarity distance", IEEE Trans. Knowl. Data Eng., vol. 19, no. 3, pp. 1065-1076, Mar. 2007.

[8] X. Qian, H. Wang, G. Liu, X. Hou, "HWVP: Hierarchical wavelet packet texture descriptors and their applications in scene categorization and semantic concept retrieval", Multimedia Tools Appl., vol. 69, no. 3, pp. 897-920, Apr. 2014.

[9] X. Qian, G. Liu, D. Guo, "Object categorization using hierarchical wavelet packet texture descriptors", Proc. IEEE 11th Int. Symp. Multimedia, pp. 44-51, 2009.

[10] X. Qian, Y. Zhao, J. Han, "Image location estimation by salient region matching", IEEE Trans. Image Process., vol. 24, no. 11, pp. 4348-4358, Nov. 2015.