Large Scale Tag Recommendation Using Different Image Representations

Rabeeh Abbasi, Marcin Grzegorzek, and Steffen Staab

ISWeb - Information Systems and Semantic Web, University of Koblenz-Landau
Universitätsstrasse 1, 56070 - Koblenz, Germany
{abbasi,marcin,staab}@uni-koblenz.de

Abstract. Nowadays, geographical coordinates (geo-tags), social annotations (tags), and low-level features are available in large image datasets. In our paper, we exploit these three kinds of image descriptions to suggest possible annotations for new images uploaded to a social tagging system. In order to compare the benefits each of these description types brings to a tag recommender system on its own, we investigate them independently of each other. First, the existing data collection is clustered separately for the geographical coordinates, tags, and low-level features. Additionally, random clustering is performed in order to provide a baseline for experimental results. Once a new image has been uploaded to the system, it is assigned to one of the clusters using either its geographical or low-level representation. Finally, the most representative tags for the resulting cluster are suggested to the user for annotation of the new image. Large-scale experiments performed for more than 400,000 images compare the different image representation techniques in terms of precision and recall in tag recommendation.

1 Introduction

With the explosive growth of Web and the recent development in digital media technology, the number of images on the Web has grown tremendously. Online photo services such as Flickr and Zooomr allow users to share their pictures with family, friends, and the online community at large. An important functionality of these services is that users manually annotate their pictures using so called *tags*, which describe their contents or provide additional contextual and semantical information. Tags are used for navigation, finding, and browsing resources and thus provide an immediate benefit for users. In practice, however, users often tag their pictures fully manually which is very time-consuming and therefore very inconvenient and expensive. For this reason, it is very important to automatize this process by developing the so called *Tag Recommender Systems* assisting users in the tagging phase.

Although this research area has been now active for a couple of years [1,2,6,7], the existing recommendation strategies are preliminary and their performance for generic scenarios rather moderate. The basic idea of recommending tags for a new image is to reuse tags assigned to similar images which have been stored in

T.-S. Chua et al. (Eds.): SAMT 2009, LNCS 5887, pp. 65–76, 2009.

the data collection before. One of the most challenging problems here is to find those similar images in a large-scale photo collection. Early approaches aimed at solving this image retrieval problem by using exclusively low-level features [13] which turned out to be almost impossible in generic environments at large-scale. Their performance was acceptable only for certain domain specific applications such as the content-based medical image retrieval [12]. Nowadays, the state-of-the-art imaging devices provide pictures together with the geographical coordinates (*geo-tags*) stating precisely where they have been acquired. Therefore, more and more researchers make use of the additional information designing tag recommender systems with quite promising results [4,8].

The most of recent tag recommendation approaches combine different image description types (geo-tags, tags, low-level features) in order to achieve reasonable results [9,11]. However, one can observe a lack of research activities comparing the benefits each of these description types brings to a tag recommender system on its own. In our paper, we exploit these three kinds of image descriptions to suggest possible annotations for new images uploaded to a collaborative tagging system independently of each other. First, we cluster the existing large-scale data collection separately for the geo-tags, tags, and low-level features. Additionally, we perform random clustering in order to provide a baseline for experimental results. Once a new image has been uploaded to the system, it is assigned to one of the clusters using either its geographical or low-level representation. Finally, the most representative tags for the resulting cluster are suggested to the user for annotation of the new image. Large-scale experiments performed for as many as $413,848$ images compare the different image representation techniques in terms of precision and recall in tag recommendation.

The paper is structured as follows. Section 2 gives an overview about our tag recommendation system. In Section 3 the content description methods (features) used in social media are shortly explained, especially those used in our framework. Section 4 explains how generating image annotations works in our framework. In Section 5, we describe the dataset used for experiments and evaluate tag recommender systems of the architecture proposed at large-scale. The tests compare different image representation techniques in terms of precision and recall in tag recommendation. Section 6 concludes our investigations and their results presented in this paper.

2 System Overview

We split the overall system for tag recommendation into two parts: training and tag recommendation. The system is trained based on the image features available in social media, once the system is trained, it is used for recommending tags for new images. Following is the brief description of training and tag recommendation phases:

Training: In the training phase images are clustered based on their features. A cluster contains homogeneous images depending upon the type of features used for clustering. For this research work, we considered geographical coordinates, low-level image features and tags as image features. As an example, a cluster based on geographical coordinates might represent the images taken in a particular location, a cluster based on low-level features might contain images showing buildings or a beach, and a cluster based on tagging data might represent concepts like concert or river. Clustering process used in this research work is described in section 4.1. Representative tags of a set of homogeneous images (i.e. images in a cluster) are used to annotate new images. The method of identifying representative tags is described in the section 4.2.

Tag Recommendation: For recommending tags to a new image, we map the image to its closest cluster and assign the representative tags of the cluster to the new image. The method of classifying an image to its closest cluster and recommending tags are described in section 4.3. In the following section, we describe the features that we have used in our experiments and are also available in Folksonomies on a large scale.

3 Features in Social Media

To analyze the effect of different type of features on the performance of tag recommendation, we use three different image features in our experiments, namely *Geographical Coordinates* (G), *Low-level image features* (L), and *Tags* (T). Following are the details of the features used in this research work.

Geographical Coordinates: With the advancement in camera and mobile technologies, nowadays many devices are available in market that are able to capture the location of the image using a built-in or external device. In addition to the possibility of capturing location of an image using a GPS device, some folksonomies like Flickr facilitate the users to add geographical coordinates to their images by providing a map interface where users can place their images on the map. Due to this easiness, there are many images in Flickr which are enriched with geographical information. In the CoPhIR dataset [3], around 4 Million out of 54 Million images are annotated with geographical coordinates. The number of geographically annotated images is supposed to increase in future as more devices will be able to capture the geographical coordinates. We represent the geographical coordinates of the images in a two dimensional vector space $G \in \Re^2$. Each row vector $\boldsymbol{g}_i$ of the feature space G represents the geographical coordinates of the image i.

Low-level Image Features: There are five different types of low-level MPEG-7 features available in the CoPhIR dataset for 54M images. Table 1 shows the properties and dimensions of the low-level features available in CoPhIR dataset. Based on initial experimental results, we consider two low-level features for evaluation, the MPEG-7 *Edge Histogram Descriptor* (EHD) and *Color Layout* (CL), which outperformed other available low-level image features. EHD represents the local edge distribution and CL represents the color and spatial information in

Table 1. Properties and dimensions of low-level features available in CoPhIR dataset

Low-level Feature	Properties	Dims
Scalable Color	Color histogram	64
Color Structure	Localized color distributions	64
Color Layout	Color and spatial information	12
Edge Histogram	Local-edge distribution	80
Homogeneous Texture	Texture	62

the images. We represent the low-level image features based on EHD and CL in 80 and 12-dimensional feature spaces $L_E \in \Re^{80}$ and $L_C \in \Re^{12}$ respectively. A row vector $\boldsymbol{\ell}_i$ of the feature space L_E or L_C represents the edge histograms or color layout of the image i respectively.

Tags: Tags are freely chosen keywords associated with the images. There is no restriction in selecting a tag for an image. A tag might represent a concept in an image, describe the image itself or it might also represent the context of the image (e.g. location, event, time etc.). On average there are only a few tags associated with the images. In 54M images of the CoPhIR dataset, each images has on average 3.1 tags. We represent the tags of the images as a n_t dimensional vector space $T \in \Re^{n_t}$, where n_t is the number of tags in the dataset. A row vector $\boldsymbol{t}_{i*}$ of the vector space T represents a resource whose non-zero values represent the tags associated with the resource i. A column vector $\boldsymbol{t}_{*j}$ represents a tag vector whose non-zero values represent the resources associated with the tag j. A value $\boldsymbol{t}_{ij}$ represents the number of times resource i is associated with the tag j.

The images in all feature spaces are indexed in the same order. For an image i, the row vector $\boldsymbol{g}_i$ represents its geographical coordinates, $\boldsymbol{l}_i$ represents its low-level image features, and $\boldsymbol{t}_i$ represents the tags associated with the same image i.

4 Tags Recommendation

This section explains the proposed tag recommendation system in detail. In the training phase of tag recommendation, the resources are first clustered (Sec. 4.1), then for each cluster, its representative tags are identified (Sec. 4.2). In the tag recommendation phase, a new resource is mapped to its closest cluster and the representative tags of the closest cluster are recommended for the new image (Sec. 4.3).

4.1 Clustering

Although many sophisticated clustering algorithms exist in literature, but the literature is still sparse for clustering high dimensional and large datasets. We use K-Means clustering algorithm in our experiments. K-Means is capable of

```
Input: Feature space $F \in \{G, L, T, D\}$, Number of clusters $k$
Output: A set of $k$ clusters
Method:
  1. Randomly select $k$ images from feature space $F$ as the initial cluster centroids
  2. Assign each image to the closest cluster
  3. Update cluster centroids
  4. If cluster centroids are changed, then repeat step 2-3
```

Fig. 1. K-Means clustering algorithm

clustering very large and high dimensional datasets. Of course, other clustering methods can also be employed in the framework, when one desires to fine tune the performances or improve the results. The K-Means algorithm we used is described in figure 1. In the following, we describe in detail how do we set different parameters for using K-Means.

Number of clusters: There is no generally accepted rule for setting the number of clusters for using K-Means. For our experiments we use the number of clusters as suggested by Mardia et al [10, page 365]. We define the number of clusters for n images as follows:

$$k = \sqrt{\frac{n}{2}} \tag{1}$$

By using k as defined in the above equation, we get same number of clusters for each feature space.

Initial Cluster Centroids: In K-Means clustering, the quality of clustering also depends on the selection of initial cluster centroids. For our experiments, k images are randomly selected. The same set of randomly selected images are used as cluster centroids for each feature space. Selecting the same set of images for different feature spaces avoids accidental improvement of one feature space over an other based on the initial centroids.

Computing distance/similarity between resources: During the clustering process, each image is assigned to its closest cluster (fig 1, step 2). We need a distance measure to compute the distance between an image and its closest centroid. The most popular distance measure used is *Euclidean Distance* [5, page 388]. Euclidean Distance between two m-dimensional vectors $\boldsymbol{f}$ and $\boldsymbol{c}$ is defined as follows:

$$euclidean(\boldsymbol{f}, \boldsymbol{c}) = \sqrt{\sum_{i=1}^{m} (\boldsymbol{f}_i - \boldsymbol{c}_i)^2} \tag{2}$$

We use euclidean distance for non-text feature spaces (i.e. geographical, low-level, and random feature spaces). For text (or tags) based feature spaces it is common to use *Cosine Similarity* [5, page 397]. We use cosine similarity to compute similarity between image tags (in feature space T) and cluster centroids. Cosine similarity between two m-dimensional vectors $\boldsymbol{f}$ and $\boldsymbol{c}$ is defined as follows:

$$cosine(\boldsymbol{f}, \boldsymbol{c}) = \frac{\boldsymbol{f}^{\mathrm{T}} \cdot \boldsymbol{c}}{||\boldsymbol{f}||||\boldsymbol{c}||} \tag{3}$$

Experimental results show that cosine similarity for tag/text based features performs significantly better than euclidean distance. For comparison between different distance measures, we also evaluated the results on Manhattan distance for non-text based features. There was no significant improvement in results if we use Manhattan distance. Manhattan distance between two vectors $\boldsymbol{f}$ and $\boldsymbol{c}$ is defined as follows

$$manhattan(\boldsymbol{f}, \boldsymbol{c}) = \sum_{i=1}^{m} |\boldsymbol{f}_i - \boldsymbol{c}_i| \tag{4}$$

4.2 Identifying Representative Tags

After clustering images into k clusters, we identify the representative tags for each cluster. The most representative tags of a cluster are recommended for the new image. To identify the representative tags of each cluster, we rank the tags by user frequency in descending order. The rank of a tag is higher if more users have used it and vice versa. We associate the top s tags to the cluster c, and represent the set of most representative top s tags associated with a cluster c as $\boldsymbol{c}_T$.

4.3 Classification and Tag Recommendation

Once we have clustered the images and identified representative tags of these clusters, we can recommend representative tags of the closest cluster from a new image. The image is mapped to the cluster, whose centroid is at minimum distance from the image. Most representative tags associated with the mapped cluster are assigned to the new image. We assume that we have the geographical coordinates and low-level features of the new image, but we do not have tags associated with the new image. In the case of clusters based on geographical or low-level feature space, we can directly measure the distance between the geographical or low-level features of the new image and the centroids of the clusters. But for tag based clustering, we do not have tags for the new image. Therefore we have to compute the centroids of tag based clusters in terms of either geographical or low-level features. For clusters based on geographical coordinates, we classify the new image to one of the clusters whose centroid is at minimum geographical distance from the new image. For low-level clusters, we classify the new image based on the distance between its low-level features and cluster centroids. For tag based clusters, as we do not have any tags for the new image, we classify the new image based on the distance between its geographical coordinates and the mean of geographical coordinates of the tag based clusters. The mismatch between feature spaces used for tag based clustering and the new image negatively effects the results of tags based clustering. To sum up the tag recommendation process, we list down the recommendation processes in three steps as follows:

1. Find closest cluster centroid $\boldsymbol{c}$ to the image $\boldsymbol{f}$ (use geographical mean as cluster centroid for geographical (G) and tag (T) based clusters; and low-level mean as cluster centroids for clusters based low-level (L) features)
2. Recommend the tags $\boldsymbol{c}_T$ associated with the cluster c to the new image

5 Experiments and Results

In this section the experiments and results are presented. The image dataset is briefly described in Section 5.1, the distinction between the training and the test data comes in Section 5.2, which is followed by the evaluation method in Section 5.3. Section 5.4 presents the comprehensive results achieved in our work.

5.1 Image Dataset

CoPhIR dataset [3] consists of images uploaded to Flickr by hundreds of thousands of different users, which makes the dataset very heterogeneous. One can find images of very different types like portraits, landscapes, people, architecture, screen shots etc. To perform an evaluation on different types of features (geo-tags, tags, low-level) on a reasonably large scale, we created a subset of the original CoPhIR dataset. We selected the images taken in national capitals[1] of all the world countries. For this purpose, we considered all the images with Euclidean distance (in terms of latitude and longitude) from center of a capital city not higher than 0.1. We ignored the capital cities which had less than $1,000$ images; this resulted into a set of 58 cities. To keep the experiments scalable, we randomly selected $30,000$ images for cities which had more than $30,000$ images. There were only three such cities *Paris*, *London*, and *Washington DC*. In the end, we had images of 58 capital cities, ranging from $1,000$ to $30,000$ images with an average of $8,000$ images per city. Total number of images in our evaluation dataset was $413,848$. For scalability, particularly for low-level image features, images are trained and evaluated separately for each city.

Base Line: In order to compare the effectiveness of different image features, we created a random feature space for the images. We assign a random value between 0 and 1 to each of the image in dataset as its random feature. We consider the random features as the baseline for comparison. Same clustering methods are applied on the random features as on the other features. Random feature space is uni-dimensional and is represented as $D \in \Re$.

[1] http://en.wikipedia.org/wiki/National_capitals

5.2 Training and Test data

It is important to carefully select the training and test datasets, because when a user uploads images to Flickr, he can perform batch operations on the set of images. For example, he can assign same tags or geographical coordinates to all the images in a batch. It is also possible that the images have similar low-level features, e.g. if the images belong to a beach or a concert. If we randomly split the images into test and training datasets, there is a chance that some images belonging to a user are used for training, while other images of the same user are used in test dataset for evaluation. Such random split may effect the final results because a test image might be mapped to a cluster containing images from the same user, having similar features as the test image. It is very likely that the test image gets annotated with perfect tags, as tags of both test and training images were provided by the same user. To make the evaluation transparent, instead of randomly splitting the resources into training and test dataset, we split the users. For each city, we use resources of 75% users for training and resources of 25% users as test dataset. No image in the test dataset is annotated by a user who has also annotated images in the training dataset. After splitting the users into training and test datasets, we use $310,590$ images for training the system and $103,258$ images used as ground truth for evaluating the system.

Another aspect of fair evaluation is the quality of the tags. There are some tags which are very common in both test and training datasets. These tags mostly represent city or country names, which can be suggested by looking into a geographical database. Some common tags might not be very specific, e. g., the tags *geotagged*, *2007*, *travel* etc. Very common tags also effect the evaluation results, as they are abundant in both test and training datasets, and are almost suggested for every test image. This results into higher precision and recall values. To make the evaluation more transparent, we do not consider the ten most frequent tags for each city and we also ignore the frequent tags *geotagged* and *geotag*, because all the images in our dataset are geo-tagged and most of the images have these two tags. For each city, we also remove the very rare tags which might be incorrectly spelled tags or tags specific to a particular user. For this reason, for each city, we ignore those tags which are used by less than three users.

5.3 Evaluation

We consider the tags associated with the $103,258$ test images as ground truth. The images in the ground truth are tagged by different users and as there is no restriction on the selection of tags for a resource, therefore the tags in ground truth are very noisy. The noise in the data leads to inferior results, but the overall results show the comparative analysis of different feature spaces. We evaluate the methods using standard evaluation methods used in information retrieval: Precision P, Recall R, and F-Measure F. The evaluation measures are defined as follows:

$$P = \frac{\mathit{Number\ of\ correctly\ suggested\ tags}}{\mathit{Number\ of\ suggested\ tags}} \tag{5}$$

$$R = \frac{\mathit{Number\ of\ correctly\ suggested\ tags}}{\mathit{Number\ of\ expected\ tags}} \tag{6}$$

$$F = \frac{2 \times P \times R}{P + R} \tag{7}$$

In addition to the standard precision and recall measures, we also computed the macro precision P_m, macro recall R_m, and macro F-Measure F_m over tags as follows:

$$P_m = \frac{\sum\limits_{t \in \mathit{Tags\ Suggested}} \frac{\mathit{\#\ of\ times\ t\ correctly\ suggested}}{\mathit{\#\ of\ times\ t\ suggested}}}{\mathit{\#\ of\ tags\ suggested}} \tag{8}$$

$$R_m = \frac{\sum\limits_{t \in \mathit{Tags\ Expected}} \frac{\mathit{\#\ of\ times\ t\ correctly\ suggested}}{\mathit{\#\ of\ times\ t\ expected}}}{\mathit{\#\ of\ tags\ expected}} \tag{9}$$

$$F_m = \frac{2 \times P_m \times R_m}{P_m + R_m} \tag{10}$$

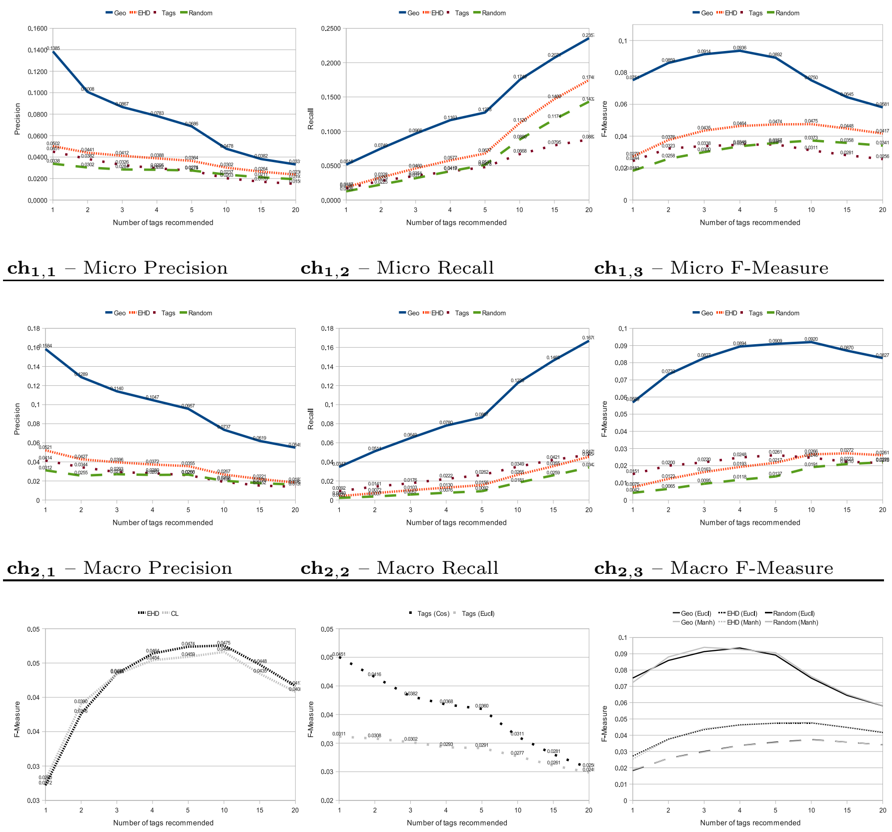
5.4 Results

The results presented in this section give a comparative view of tag recommendation based on different types of features. The automated evaluation on one hand provides the possibility to do evaluation on a large scale, but on the other hand the ground truth (test data) might contain invalid tags. We try to make the evaluation transparent and more meaningful by filtering certain types of tags (see Section 5.2). By removing very common tags, there is a certain decrease in evaluation results, but we believe that filtering make the evaluation fair. We have also evaluated the results without filtering the dataset, and in that case even random feature space gives a F-Measure value of 0.42. This is because of the reason that very common tags are recommended for the test images and there is always a major overlap between common tags of training and test data. The precision, recall, and F-Measure values presented in this section might appear to be low for the reader, but one shall keep in mind the filtering applied on the dataset to make the evaluation transparent.

Table 2 consists of nine charts ($\mathrm{ch}_{i\in\{1,2,3\},j\in\{1,2,3\}}$) presenting the experimental results.

Charts in the first row ($\mathrm{ch}_{1,j\in\{1,2,3\}}$) depict the so called *micro average* evaluation and were generated in accordance to the evaluation criteria (5), (6), and (7)

Table 2. Result charts ($\mathbf{ch}_{\mathbf{i}\in\{\mathbf{1},\mathbf{2},\mathbf{3}\},\mathbf{j}\in\{\mathbf{1},\mathbf{2},\mathbf{3}\}}$)



$\mathbf{ch}_{\mathbf{1,1}}$ – Micro Precision

$\mathbf{ch}_{\mathbf{1,2}}$ – Micro Recall

$\mathbf{ch}_{\mathbf{1,3}}$ – Micro F-Measure

$\mathbf{ch}_{\mathbf{2,1}}$ – Macro Precision

$\mathbf{ch}_{\mathbf{2,2}}$ – Macro Recall

$\mathbf{ch}_{\mathbf{2,3}}$ – Macro F-Measure

$\mathbf{ch}_{\mathbf{3,1}}$ – Micro F-Measure comparing results of two different low-level features Edge Histogram Descriptor (EHD) and Color Layout (CL)

$\mathbf{ch}_{\mathbf{3,2}}$ – Micro F-Measure comparison of Cosine (Cos) and Euclidean (Eucl) distances for tag/text based features

$\mathbf{ch}_{\mathbf{3,3}}$ – Micro F-Measure comparison of Manhattan (Manh) and Euclidean (Eucl) distances for non-text based features. Dark lines show the results obtained using Manhattan distance and gray lines show results obtained using euclidean distance

respectively. As one can see, in all three cases the results are significantly better when using geo-tags for image description. The performance of the tag recommendation using low-level features and textual tags differs only slightly from the results based on random clustering. For exactly one tag being recommended, the precision amounts to: 0.1385 for geo-tags, 0.0502 for low-level features, 0.0451 for textual tags, and 0.0338 for random clustering.

Charts in the second row ($\mathrm{ch}_{2,j\in\{1,2,3\}}$) present the so called *macro average over tags* evaluation and were generated in accordance to the evaluation criteria

(8), (9), and (10) respectively. Similar to the micro average evaluation, the results here are significantly better for geo-tags, while the performance in case of textual tags, low-level features, and random clustering is almost the same. For exactly one tag being recommended, the macro precision for geo-tags amounts to 0.1584, for low-level features - 0.0521, for textual tags - 0.0414, and the baseline is 0.0312.

In the third row of charts ($\mathrm{ch}_{3,j\in\{1,2,3\}}$) in Table 2 some further evaluations can be found. In the first and second row charts ($\mathrm{ch}_{i\in\{1,2\},j\in\{1,2,3\}}$) the Edge Histogram Descriptor (EHD) was applied whenever low-level features were used and cosine similarity was used for tag based feature space. This has got an experimental reason. As you can see in Chart ($\mathrm{ch}_{3,1}$), the EHD performs slightly better than the Color Layout (CL) in terms of micro F-Measure and chart ($\mathrm{ch}_{3,2}$) shows a clear advantage of the cosine distance over the euclidean distance for tag based features. And finally, Chart ($\mathrm{ch}_{3,3}$) explains why using the simple Euclidean distance has appeared to be sufficient in our approach. The results remain almost the same when using Manhattan distance.

6 Conclusion

In our paper, we exploited three kinds of image description techniques, namely geo-tags, tags, and low-level features, to suggest possible annotations for new images uploaded to a social tagging system. In order to compare the benefits each of these description types brings to a tag recommender system on its own, we investigated them independently of each other. The evaluation was done on a large-scale image database. For experiments we used the CoPhIR dataset [3] including images uploaded to Flickr by hundreds of thousands of different users. The processing chain of our algorithm for generating image annotations contains: (i) clustering the images, (ii) finding representative tags for the clusters, (iii) classification of new images and tag recommendation. The results showed that geo-tags are the most helpful image descriptors for tag recommendation, while textual tags and low-level features provide only a slightly better performance than the random baseline.

In the future, we will keep investigating the tag recommendation problem for large-scale heterogeneous image archives. We will further develop our framework to allow comprehensive experimental studies. We will also investigate the problem for some more domain dependent data collections.

Acknowledgments

This work has been partially supported by the European project *Semiotic Dynamics in Online Social Communities* (Tagora, FP6-2005-34721). We would also like to acknowledge Higher Education Commission of Pakistan (HEC) and German Academic Exchange Service (DAAD) for providing scholarship and support to Rabeeh Abbasi for conducting his PhD.

References

1. Adrian, B., Sauermann, L., Roth-Berghofer, T.: Contag: A semantic tag recommendation system. In: Pellegrini, T., Schaffert, S. (eds.) Proceedings of I-Semantics 2007, September 2007, pp. 297–304. JUCS (2007)
2. Basile, P., Gendarmi, D., Lanubile, F., Semeraro, G.: Recommending smart tags in a social bookmarking system. In: Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007), pp. 22–29 (2007)
3. Bolettieri, P., Esuli, A., Falchi, F., Lucchese, C., Perego, R., Piccioli, T., Rabitti, F.: CoPhIR: a test collection for content-based image retrieval. CoRR, abs/0905.4627v2 (2009)
4. Cristani, M., Perina, A., Castellani, U., Murino, V.: Content visualization and management of geo-located image databases. In: CHI 2008: CHI 2008 extended abstracts on Human factors in computing systems, pp. 2823–2828. ACM, New York (2008)
5. Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 2nd edn. Morgan Kaufmann, San Francisco (2006)
6. Heymann, P., Ramage, D., Garcia-Molina, H.: Social tag prediction. In: SIGIR 2008: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, pp. 531–538. ACM, New York (2008)
7. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in folksonomies. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) PKDD 2007. LNCS (LNAI), vol. 4702, pp. 506–514. Springer, Heidelberg (2007)
8. Kennedy, L., Naaman, M., Ahern, S., Nair, R., Rattenbury, T.: How flickr helps us make sense of the world: context and content in community-contributed media collections. In: MULTIMEDIA 2007: Proceedings of the 15th international conference on Multimedia, pp. 631–640. ACM, New York (2007)
9. Kennedy, L.S., Naaman, M.: Generating diverse and representative image search results for landmarks. In: WWW 2008: Proceeding of the 17th international conference on World Wide Web, pp. 297–306. ACM, New York (2008)
10. Mardia, K., Kent, J., Bibby, J.: Multivariate Analysis. Academic Press, London (1979)
11. Moëllic, P.-A., Haugeard, J.-E., Pitel, G.: Image clustering based on a shared nearest neighbors approach for tagged collections. In: CIVR 2008: Proceedings of the 2008 international conference on Content-based image and video retrieval, pp. 269–278. ACM, New York (2008)
12. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medical applications - clinical benefits and future directions. International Journal of Medical Informatics 73(1), 1–23 (2003)
13. Pentland, A., Picard, R., Sclaroff, S.: Tools for content-based manipulation of image databases. International Journal of Computer Vision 18(3), 233–254 (1996)