

SIMULTANEOUS IMAGE TAGGING AND GEO-LOCATION PREDICTION WITHIN HYPERGRAPH RANKING FRAMEWORK

Konstantinos Pliakos and Constantine Kotropoulos

Department of Informatics, Aristotle University of Thessaloniki
Box 451, Thessaloniki, 54124, Greece
Email: {kpliakos, costas}@aiaa.csd.auth.gr

ABSTRACT

The development of social media has led to a burst of interest in image-related metadata information, such as tags and geo-tags. Tags are semantic keywords that are assigned to an image. Image tagging enables the users of social media sharing platforms to annotate images, facilitating image search and content description. Despite the volume of related research, issues such as accuracy or efficiency still remain open problems. Here, a novel method for simultaneous image tagging and geo-location prediction is proposed that is based on hypergraph learning. The method is further improved by enforcing group sparsity constraints. It fully exploits various types of information, such as social, image-related metadata, or similarities based on visual attributes. Experiments on a dataset crawled from *Flickr* demonstrate F_1 at 10 top ranked tags equal to 0.558 for image tagging and cumulative geotagging prediction rate at 3 top ranks equal to 83%.

Index Terms— Tagging, Recommender systems, Hypergraph, Group Sparsity Optimization, Geo-coordinate prediction

1. INTRODUCTION

Nowadays, many social media sharing platforms have been developed and have gained significant popularity. Various websites like *Flickr*¹ or *Picasa Web Album*² enable users to annotate images, describing their content according to the users' point of view. The perception of image content by humans may differ, for example, due to cultural differences among the people. Image tagging aims at bridging this semantic gap, facilitating image search and categorization. This task is further enhanced by exploiting another image-related type of metadata information, the geo-tags. Modern mobile devices like cameras or smart-phones have the ability to assign specific geo-coordinates to an image by the time it is taken automatically. Enriching the image with this valuable geographic information leverages image search. Clearly, the exponential growth of the number of uploaded images on the web exemplifies the need for efficient and accurate tagging.

In the past, many works focused on image tagging, exploiting the image content. S. L. Feng *et al.* [1] proposed an image and video annotation model using the joint probability distribution of the possible annotations and the image feature vectors. The estimation of the annotation probabilities was based on a multiple Bernoulli model, while non-parametric kernel density estimates were used as image features. X. Zhang *et al.* [2] proposed a probabilistic model based

on Markov Random Fields in order to capture the correlations between the different features. Information from multi-type interrelated objects was exploited in [3,4], using simple graph-based methods. Another image tagging method was presented in [5], where a joint classification and regression model was used to perform image annotation and geo-tag prediction, simultaneously. A different solution to the geo-tag prediction problem was based on a data-driven scene matching approach [6]. Hypergraph based approaches were proposed for 3-D object retrieval [7] and for estimating the relevance of user tagged images [8]. In [9], the spatial group sparse coding was proposed for localizing tags to image regions, extending group sparse coding with spatial correlations among the regions.

Here, a novel approach to simultaneous image tagging and geo-location prediction within a hypergraph ranking framework is proposed. Hypergraphs consist of a set of vertices made by concatenating different kinds of objects (images, users, social groups, geo-tags, tags) and hyperedges linking these vertices [10–15]. This way, the existing multi-link relations between the vertices are represented. By fully exploiting the information distilled from social media, hypergraphs are demonstrated to outperform existing methods that are based on visual image features only or graph-based methods, which model only the pairwise relations between the images [2,6]. Indeed, the context revealed by user-ratings, such as tags and geo-tags, is found to be more crucial than the content descriptive features [16]. Such findings by no means de-emphasize the role of content analysis, but undoubtedly underline the need for more powerful feature analyzers. Here, it is demonstrated that the inclusion of user friendship and user group relations on the top of image-related metadata and image similarity increases the accuracy of both image tagging and geo-location prediction. Building on hypergraph ranking, the latter is enhanced by enforcing group sparsity constraints. This way, the set of objects (i.e., vertices) is split into different object groups (images, users, social groups, tags, geo-tags) and each object group effect in image tagging and geo-location prediction is controlled separately by assigning them different weights. Experiments on a dataset crawled from *Flickr* demonstrate F_1 at 10 top ranked tags equal to 0.558 for image tagging and cumulative geotagging prediction rate at 3 top ranks equal to 83%.

The remainder of this paper is organized as follows. In Section 2, the ranking on a hypergraph, enforcing group sparsity, is detailed. In Section 3, the dataset is described and the hypergraph construction is explained. Experimental results are presented in Section 4, demonstrating the effectiveness of the proposed method. Conclusions are drawn and topics for future research are indicated in Section 5.

¹<http://www.flickr.com>

²<http://picasaweb.google.com>

2. HYPERGRAPH RANKING WITH GROUP SPARSE REGULARIZATION

In the following, $|\cdot|$ denotes set cardinality, $\|\cdot\|_2$ is the ℓ_2 norm of a vector, and \mathbf{I} is the identity matrix of compatible dimensions. A hypergraph is a generalization of a graph having edges, which connect more than two vertices. Let $G(V, E, w)$ denote a hypergraph with set of vertices V and set of hyperedges E to which a weight function $w: E \rightarrow \mathbb{R}$ is assigned. The vertex set V is made by concatenating sets of objects of different type (images, users, social groups, geo-tags, tags). These vertices and hyperedges form a $|V| \times |E|$ incidence matrix with elements $H(v, e) = 1$ if $v \in e$ and 0 otherwise. Based on \mathbf{H} , the vertex and hyperedge degrees are defined as $\delta(v) = \sum_{e \in E} w(e)H(v, e)$ and $\delta(e) = \sum_{v \in V} H(v, e)$, respectively. The following diagonal matrices are defined: the vertex degree matrix \mathbf{D}_v of size $|V| \times |V|$, the hyperedge degree matrix \mathbf{D}_e of size $|E| \times |E|$, and the $|E| \times |E|$ matrix \mathbf{W} containing the hyperedge weights.

Let $\mathbf{A} = \mathbf{D}_v^{-1/2} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^T \mathbf{D}_v^{-1/2}$. Then, $\mathbf{L} = \mathbf{I} - \mathbf{A}$ is the positive semi-definite Laplacian matrix of the hypergraph. The elements of \mathbf{A} , $A(u, v)$, indicate the relatedness between the objects u and v . A real valued ranking vector $\mathbf{f} \in \mathbb{R}^{|V|}$ that minimizes $\Omega(\mathbf{f}) = \frac{1}{2} \mathbf{f}^T \mathbf{L} \mathbf{f}$ yields a clustering of the hypergraph, where all vertices with the same value in the ranking vector \mathbf{f} are strongly connected [17]. By including the ℓ_2 regularization norm between the ranking vector \mathbf{f} and a query vector $\mathbf{y} \in \mathbb{R}^{|V|}$, a recommendation problem can be solved [15]. That is, the function to be minimized is expressed as

$$\tilde{\Psi}(\mathbf{f}) = \Omega(\mathbf{f}) + \vartheta \|\mathbf{f} - \mathbf{y}\|_2^2 \quad (1)$$

where ϑ is a regularization parameter. The ranking vector $\mathbf{f}^* = \arg \min_{\mathbf{f}} \tilde{\Psi}(\mathbf{f})$ is [15]:

$$\mathbf{f}^* = \frac{\vartheta}{1 + \vartheta} \left(\mathbf{I} - \frac{1}{1 + \vartheta} \mathbf{A} \right)^{-1} \mathbf{y}. \quad (2)$$

Hereafter, each vertex subset is referred to as *object group* in order to avoid confusion with social groups. Indisputably, each object group contributes differently to the ranking procedure. A Group Lasso regularizing term is more appropriate than the ℓ_2 norm in this kind of problems [18]. The hypergraph vertices are split into S non-overlapping object groups (images, users, social groups, geo-tags, tags) and different weights γ_s , $s = 1, 2, \dots, S$ are assigned to each object group, yielding the following objective function to be minimized:

$$\Psi(\mathbf{f}) = \Omega(\mathbf{f}) + \vartheta \sum_{s=1}^S \sqrt{\gamma_s (\mathbf{f} - \mathbf{y})^T \mathbf{K}_s (\mathbf{f} - \mathbf{y})}. \quad (3)$$

In (3), ϑ is also a regularization parameter and \mathbf{K}_s is the $|V| \times |V|$ diagonal matrix with elements equal to 1 for the vertices, which belong to the s -th object group. The latter minimization problem is expressed as:

$$\mathbf{f}^* = \arg \min_{\mathbf{f}} \Psi(\mathbf{f}). \quad (4)$$

Let $\mathbf{x} = \mathbf{f} - \mathbf{y}$. By introducing the auxiliary variable $\mathbf{z} = \mathbf{x}$, the right-hand side of (4) is rewritten as:

$$\arg \min_{\mathbf{x}} \frac{1}{2} (\mathbf{x} + \mathbf{y})^T \mathbf{L} (\mathbf{x} + \mathbf{y}) + \vartheta \sum_{s=1}^S \sqrt{\gamma_s \mathbf{z}^T \mathbf{K}_s \mathbf{z}} \quad (5)$$

s.t. $\mathbf{z} = \mathbf{x}$.

The solution of (5) can be obtained by minimizing the augmented Lagrangian function

$$\mathcal{L}(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) = \frac{1}{2} (\mathbf{x} + \mathbf{y})^T \mathbf{L} (\mathbf{x} + \mathbf{y}) + \vartheta \sum_{s=1}^S \sqrt{\gamma_s \mathbf{z}^T \mathbf{K}_s \mathbf{z}} + \boldsymbol{\lambda}^T (\mathbf{z} - \mathbf{x}) + \frac{\mu}{2} \|\mathbf{z} - \mathbf{x}\|_2^2, \quad (6)$$

where $\boldsymbol{\lambda}$ is the vector of the Lagrange multipliers, which is updated

Algorithm 1 Alternating Directions Method

- 1: Given \mathbf{x}^t , \mathbf{z}^t and $\boldsymbol{\lambda}^t$.
 - 2: Set tolerance ϵ and initialize μ^0 .
 - 3: $\mathbf{x}^{t+1} \leftarrow \arg \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mathbf{z}^t, \boldsymbol{\lambda}^t)$
 - 4: $\mathbf{z}^{t+1} \leftarrow \arg \min_{\mathbf{z}} \mathcal{L}(\mathbf{x}^{t+1}, \mathbf{z}, \boldsymbol{\lambda}^t)$
 - 5: **if** $\|\mathbf{z} - \mathbf{x}\|_2^2 > \epsilon$ **then**
 - 6: $\boldsymbol{\lambda}^{t+1} \leftarrow \boldsymbol{\lambda}^t + \mu^t (\mathbf{z}^{t+1} - \mathbf{x}^{t+1})$
 - 7: $\mu^{t+1} = \min(1.1\mu^t, 10^6)$
 - 8: **else**
 - 9: return \mathbf{x}^{t+1} , \mathbf{z}^{t+1} .
 - 10: $\mathbf{f} = \mathbf{x}^{t+1} + \mathbf{y}$
 - 11: **end if**
-

at each iteration and μ is a parameter regularizing the violation of the constraint $\mathbf{x} = \mathbf{z}$. (6) can be solved by the Alternating Directions Method [19], as shown in Algorithm 1. Solving for \mathbf{x}^{t+1} in line 3 yields

$$\mathbf{x}^{t+1} = (\mathbf{L} + \mu^t \mathbf{I})^{-1} (\boldsymbol{\lambda}^t + \mu^t \mathbf{z}^t - \mathbf{L} \mathbf{y}). \quad (7)$$

A careful look at (7) reveals that matrix inversion is not needed at each iteration. Only one eigen-decomposition is needed. Indeed, let $\mathbf{Q}_t = \mathbf{L} + \mu^t \mathbf{I}$. Then, $\mathbf{Q}_t^{-1} = \frac{1}{\mu^t - \mu^{t-1}} \left[\mathbf{I} + \frac{1}{\mu^t - \mu^{t-1}} \mathbf{Q}_{t-1} \right]^{-1}$. $\mathbf{Q}_0 = \mathbf{L} + \mu^0 \mathbf{I}$ is a symmetric matrix. Therefore, it is diagonalizable: $\mathbf{Q}_0 = \mathbf{U} \boldsymbol{\Lambda}_0 \mathbf{U}^T$, where $\mathbf{U} \mathbf{U}^T = \mathbf{U}^T \mathbf{U} = \mathbf{I}$. It can be easily derived that $\mathbf{Q}_1^{-1} = \mathbf{U} [(\mu^1 - \mu^0) \mathbf{I} + \boldsymbol{\Lambda}_0]^{-1} \mathbf{U}^T$ and in general

$$\mathbf{Q}_t^{-1} = \mathbf{U} [(\mu^t - \mu^0) \mathbf{I} + \boldsymbol{\Lambda}_0]^{-1} \mathbf{U}^T. \quad (8)$$

The minimization problem described in line 4 of Algorithm 1 is expressed as

$$\min_{\mathbf{z}} \mu^t \left\{ \vartheta \sum_{s=1}^S \sqrt{\gamma_s \mathbf{z}^T \mathbf{K}_s \mathbf{z}} + \frac{1}{2} \|\mathbf{z} - (\mathbf{x}^{t+1} - \frac{1}{\mu^t} \boldsymbol{\lambda}^t)\|_2^2 \right\}. \quad (9)$$

By applying the soft-thresholding operator [20], we obtain

$$z_j = \frac{r_j}{\|\mathbf{r}_s\|_2} \max \left(0, \|\mathbf{r}_s\|_2 - \vartheta \mu^t \frac{1}{\sqrt{\gamma_s}} \right) \quad (10)$$

where $r_j = x_j^{t+1} - \frac{1}{\mu^t} \lambda_j^t$, s is the object group where the j -th element belongs, and \mathbf{r}_s denotes the segment of \mathbf{r} corresponding to the s -th object group.

3. DATASET DESCRIPTION AND HYPERGRAPH CONSTRUCTION

3.1. Dataset description

For evaluation purposes, an image dataset was collected from *Flickr*. It contains both indoor and outdoor medium sized photos of popular Greek landmarks, various city scenes, and landscapes. Using *FlickrApi*³, a large set of “geotagged” images was down-

³<http://www.flickr.com/services/api>

Table 1. Dataset objects, notations, and counts.

Object	Notation	Count
Images	Im	1292
Users	U	440
User Groups	Gr	1644
Geo-tags	Geo	125
Tags	Ta	2366

loaded along with valuable information related to them (id, title, owner, latitude, longitude, tags, image views). Then, the dataset was filtered based on image views (times that the specific image has been seen in *Flickr*) and owner’s uploading statistics. At this point, it was assumed that images with many views normally depict worth seeing landmarks and owners (users) with many uploaded images were active ones, possessing many social relations (friends, social groups). The owners of these images were the users in the dataset. Then, corresponding social information (friends, social groups) was crawled and only the groups that had at least 5 owners from the dataset as members were kept. The specific cardinalities are summarized in Table 1.

In order to form a proper set of tags, all characters were converted to lower case, unreadable symbols and redundant information were removed. Next, a vocabulary of unique words was generated along with their frequencies. Then, terms with frequency 1 or 2 were removed from the set of tags and the vocabulary. Finally, spelling mistakes were corrected and any morphological variations merged using the Edit Distance [21].

Geo-tags were clustered into 125 different clusters using hierarchical clustering, having computed pairwise distances according to the “Haversine formula”⁴.

3.2. Hypergraph construction

The vertex set is defined as $V = Im \cup U \cup Gr \cup Geo \cup Ta$. The hypergraph, \mathbf{H} is formed concatenating the 6 hyperedge sets indicated as columns in Table 2. \mathbf{H} has a size of 5867×30924 elements. In the following, the weights of the hyperedge sets $E^{(1)}-E^{(5)}$ are set equal to one. The dataset has captured 2276 friendship relations and 19127 tagging ones.

$E^{(1)}$ represents a pairwise friendship relation between users. The incidence matrix of the hypergraph $UE^{(1)}$ has a size of 440×2276 elements.

$E^{(2)}$ represents a user group. It contains all the vertices of the corresponding users as well as the ones corresponding to the user group. The incidence matrix of the hypergraph $UE^{(2)} - GrE^{(2)}$ has a size of 2084×1644 elements.

$E^{(3)}$ contains a user and an uploaded image, representing a user-image possession relation. Each image has only one owner. The incidence matrix of the hypergraph $UE^{(3)} - ImE^{(3)}$ has a size of 1732×1292 elements.

$E^{(4)}$ captures a geo-location relation. This hyperedge set contains triplets of Im , U and Geo . The incidence matrix of the hypergraph $ImE^{(4)} - UE^{(4)} - GeoE^{(4)}$ has a size of 1857×125 elements.

$E^{(5)}$ also contains triplets, Im , U and Ta . Each hyperedge represents a tagging relation. The incidence matrix of the hypergraph $ImE^{(5)} - UE^{(5)} - TaE^{(5)}$ has a size of 4098×19127 elements.

⁴<http://www.movable-type.co.uk/scripts/latlong.html>

Table 2. The structure of the hypergraph incidence matrix \mathbf{H} and its sub-matrices.

$E^{(1)}$	$E^{(2)}$	$E^{(3)}$	$E^{(4)}$	$E^{(5)}$	$E^{(6)}$
0	0	$ImE^{(3)}$	$ImE^{(4)}$	$ImE^{(5)}$	$ImE^{(6)}$
$UE^{(1)}$	$UE^{(2)}$	$UE^{(3)}$	$UE^{(4)}$	$UE^{(5)}$	0
0	$GrE^{(2)}$	0	0	0	0
0	0	0	$GeoE^{(4)}$	0	0
0	0	0	0	$TaE^{(5)}$	0

$E^{(6)}$ contains pairs of vertices, which represent two images. The weight $w(e_{ij}^{(6)})$ is set as the similarity between images i and j normalized to eliminate the bias: $w(e_{ij}^{(6)})' = \frac{w(e_{ij}^{(6)})}{\max(w(e_{ij}^{(6)}))}$. In order

to form this part of the hypergraph, both global and local features were used. Firstly, the 100 nearest neighbors to each image were identified using the GIST descriptors [22] and they were reduced to the 5 most similar images to the reference image, by using scale-invariant feature transform (SIFT) [23]. The incidence matrix of the hypergraph $ImE^{(6)}$ has a size of 1292×6460 .

The query vector \mathbf{y} is initialized by setting the entry corresponding to the reference image im and its owner u to 1. The tags ta connected to this image are set to $A(im, ta)$. The objects corresponding to gr and geo associated to the image owner u are set to $A(u, gr)$ and $A(u, geo)$, respectively. The query vector \mathbf{y} has a length of 5867 elements.

The ranking vector \mathbf{f}^* is derived by solving either (2) or (4), having set the query vector \mathbf{y} , the regularization parameter ϑ , and the group of objects weights γ_s in case of (4), as detailed in Section 4. It has the same size and structure as \mathbf{y} . The values corresponding to tags are used for image tagging with the top ranked tags being recommended for the reference image. The values corresponding to geo are used for geo-location prediction with only the 3 top ranked geo-locations (i.e., geo-clusters) being recommended for the reference image, consecutively.

4. EXPERIMENTS

The averaged Recall-Precision and F_1 measure are used as figures of merit. Precision is defined as the number of correctly recommended tags divided by the number of all recommended tags. Recall is defined as the number of correctly recommended tags divided by the number of all tags the user has actually set. The F_1 measure is the weighted harmonic mean of precision and recall, which measures the effectiveness of recommendation when treating precision and recall as equally important. The ranking obtained by (2) is referred to as Image Tagging on Hypergraph (ITH), while that obtained by (4) is coined as Query Group Sparse Optimization (QGSO). The geo-location prediction is referred to as GPR.

For evaluation purposes, a test set containing the 25% of the tags and a training set containing the remaining 75% are defined. During testing, the tags contained in the test set were not included in the training procedure. That is, the associated elements to the test image are set equal to zero in \mathbf{A} and \mathbf{y} . The relations between test images im and geo-locations (i.e., geo-clusters geo) are set also to 0. The results of the (ITH) are demonstrated in Fig. 1, in which the averaged Recall-Precision curves are plotted by averaging the Recall-Precision curves over 1186 images with at least 4 tags. To calculate the recall and precision, the 15 top ranked tags are being

recommended to any test image. For the GPR, only the 3 top ranked elements are taken from the part of the f vector associated to the geo-locations. In Fig. 2, the results for the GPR are presented.

By enforcing group sparsity in the ranking problem, the performance is improved significantly, as shown in Fig. 1. The weights for the 5 different object groups (images, users, user groups, geo-tags and tags) were set to 0.9, 0.9, 0.6, 0.2, and 0.2, respectively. This choice was made empirically. The typical values of μ^0 , ϵ , and ϑ are 10^{-6} , 10^{-8} , and 2, respectively.

In Table 3, the averaged F_1 measure is listed for the ITH and the QGSO, corresponding to 5 different ranking positions. It is evident that both methods achieve an encouraging F_1 measure. It is clearly indicated that the QGSO outperforms the ITH.

In Table 4, the rate of correct geo-location prediction is detailed for both the ITH and the QGSO, corresponding to the 3 top ranked elements. Taking strictly only the 1-st element, a correct prediction rate equal to 56% is obtained with the ITH and by taking the 3 top ranked elements, the rate reaches 81%. The QGSO method demonstrates higher rates than the ITH by 2%. These results are further compared with those obtained by exploiting geographical information deduced from the tags. Greek geo-names were collected from the GeoNames geographical database⁵ along with their geo-coordinates. A geo-location prediction was made for each image having tags matching the geo-names. The distance between the image geo-coordinates, treated as ground truth, and the ones associated to the geo-name was computed by using ‘‘Haversine formula’’. Predictions whose distances fall below a threshold of 500 m were considered as correct. Let us call, the just described naive approach as GNPR. Fig.2 depicts the geo-location prediction rates achieved by the ITH, the QGSO, and the GNPR. Clearly, the QGSO yields the highest rate. GNPR yields a correct prediction rate equal to 38%.

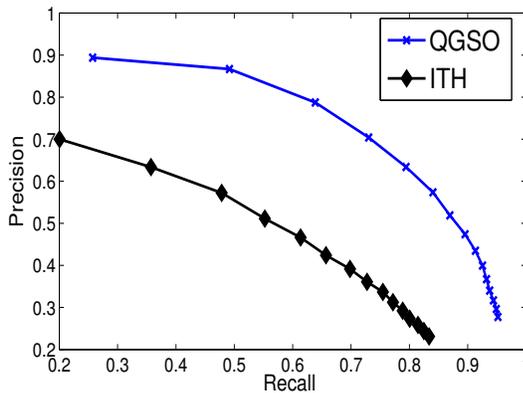


Fig. 1. Averaged Recall-Precision curves for the ITH and the QGSO.

Table 3. F_1 measure for the ITH and the QGSO at ranking positions 1, 2, 5, 10, and 15.

	$F_1@1$	$F_1@2$	$F_1@5$	$F_1@10$	$F_1@15$
ITH	0.312	0.457	0.530	0.445	0.362
QGSO	0.400	0.627	0.705	0.558	0.430

⁵<http://www.geonames.org>

Table 4. GPR (in %) for the ITH and the QGSO at 3 ranking positions and cumulative rank score.

	1st rank	2nd rank	3rd rank	cumulative
ITH	56%	17%	8%	81%
QGSO	57%	18%	8%	83%

5. CONCLUSION AND FUTURE WORK

In this paper, a novel and efficient method of simultaneous image tagging and geo-location prediction has been proposed. It fully exploits the image content, the context, and the social media information. Thanks to hypergraph learning, image tagging and geo-location prediction have been addressed and further improved by enforcing group sparsity constraints. The merits of the proposed method were demonstrated experimentally on a collection of Greek landmark images. The method can also accommodate tagging for any multimedia (e.g., music, video) or even the fusion between them. Applications, such as travel guideline systems based on images and their geographic coordinates or query-based systems recommending touristic destinations, can benefit from the proposed method.

6. ACKNOWLEDGMENTS

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operation Program ‘‘Competitiveness-Cooperation 2011’’ - Research Funding Program: SYN-10-1730-ATLAS.

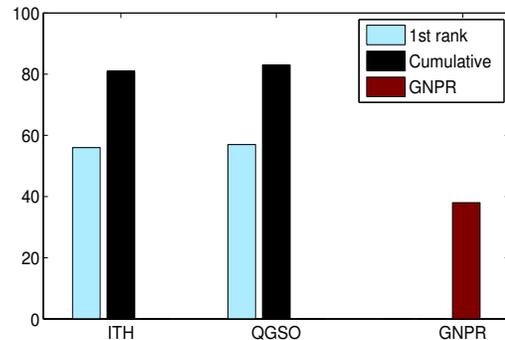


Fig. 2. Geo-location prediction rates (in %) for the compared methods.

7. REFERENCES

- [1] S. L. Feng, R. Manmatha, and V. Lavrenko, "Multiple Bernoulli relevance models for image and video annotation," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 2004, vol. 2, pp. II-1002-II-1009.
- [2] X. Zhang, Z. Li, and W. Chao, "Tagging image by merging multiple features in a integrated manner," *Journal of Intelligent Information Systems*, vol. 39, no. 1, pp. 87-107, 2012.
- [3] X. Zhang, X. Zhao, Z. Li, J. Xia, R. Jain, and W. Chao, "Social image tagging using graph-based reinforcement on multi-type interrelated objects," *Signal Processing*, vol. 93, no. 8, pp. 2178-2189, 2013.
- [4] Z. Guan, J. Bu, Q. Mei, C. Chen, and C. Wang, "Personalized tag recommendation using graph-based ranking on multi-type interrelated objects," in *Proc. 32nd Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, 2009, pp. 540-547.
- [5] H. Wang, D. Joshi, J. Luo, H. Huang, and M. Park, "Simultaneous image annotation and geo-tag prediction via correlation guided multi-task learning," in *Proc. IEEE Int. Symp. Multimedia*, 2012, pp. 69-72.
- [6] J. Hays and A. A. Efros, "IM2GPS: Estimating geographic information from a single image," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1-8.
- [7] Y. Gao, M. Wang, D. Tao, R. Ji, and Q. Dai, "3-d object retrieval and recognition with hypergraph analysis," *IEEE Trans. Image Processing*, vol. 21, no. 9, pp. 4290-4303, 2012.
- [8] Y. Gao, M. Wang, Z. Zha, J. Shen, X. Li, and X. Wu, "Visual-textual joint relevance learning for tag-based social image search," *IEEE Trans. Image Processing*, vol. 22, no. 1, pp. 363-376, 2013.
- [9] Y. Yang, Y. Yang, Z. Huang, H. T. Shen, and F. Nie, "Tag localization with spatial correlations and joint group sparsity," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2011, pp. 881-888.
- [10] C. Berge and Edward Minieka, *Graphs and Hypergraphs*, vol. 7, North-Holland, Amsterdam, 1973.
- [11] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Advances in Neural Information Processing Systems*, 2006, pp. 1601-1608.
- [12] J. Yu, D. Tao, and M. Wang, "Adaptive hypergraph learning and its application in image classification," *IEEE Trans. Image Processing*, vol. 21, no. 7, pp. 3262-3272, 2012.
- [13] Z. Yu, S. Tang, Y. Zhang, and J. Shao, "Image ranking via attribute boosted hypergraph," in *Proc. 13th Pacific-Rim Conf. Advances in Multimedia Information Processing*, 2012, pp. 779-789.
- [14] J. Xu, V. Singh, Z. Guan, and B. S. Manjunath, "Unified hypergraph for image ranking in a multimodal context," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2012, pp. 2333-2336.
- [15] J. Bu, S. Tan, C. Chen, C. Wang, H. Wu, Z. Lijun, and X. He, "Music recommendation by unified hypergraph: Combining social media information and music content," in *Proc. ACM Conf. Multimedia*, 2010, pp. 391-400.
- [16] M. Slaney, "Web-scale multimedia analysis: Does content matter?," *IEEE Multimedia*, vol. 11, no. 2, pp. 12-15, April-June 2011.
- [17] S. Agarwal, K. Branson, and S. Belongie, "Higher order learning with graphs," in *Proc. 23rd Int. Conf. Machine Learning*, 2006, pp. 17-24.
- [18] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49-67, 2006.
- [19] Z. Lin, R. Lui, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Advances Neural Information Processing Systems*, 2011, pp. 612-620.
- [20] Z. Qin and K. Scheinberg, "Efficient block-coordinate descent algorithms for the Group Lasso," *Industrial Engineering*, pp. 1-21, 2010.
- [21] E. S. Ristad and P. Yianilos, "Learning string-edit distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 5, pp. 522-532, 1998.
- [22] A. Oliva and A. Torralba, "Building the GIST of a scene: The role of global image features in recognition," *Progress in Brain Research*, vol. 155, pp. 23-36, 2006.
- [23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.