Boosted Inductive Matrix Completion for Image Tagging

Yuqing Hou(✉)

Key Laboratory of Machine Perception (MOE), School of EECS, Peking University, Beijing 100871, China
houyuqing1988@gmail.com

Abstract. Search engines have traditionally used manual image tagging for indexing and retrieving image collections. Manual tagging is expensive and labor intensive, motivating the research on automatic tag completion. However, existing tag completion approaches suffer from deficient or inaccurate tags. In this study, we formulate the task in the boosted inductive matrix completion (BIMC) framework, which combines the power of the inductive matrix completion (IMC) model together with a standard matrix completion (MC) model. We incorporates visual-tag correlation and semantic-tag correlation properties into the model for better exploration of the latent connection between image features and tags. We exploit CNN features and word vectors to narrow the semantic gap. The proposed method achieves good performance on several benchmark datasets with missing and noisy tags.

Keywords: Image tag completion · Boosted inductive matrix completion · Visual-tag correlation · Semantic-tag correlation · CNN features · Word vectors

1 Introduction and Motivation

Many machine learning methods have been developed for the image tag completion task. However, most methods are usually region-based, depending heavily on the image segmentation accuracy. In recent years, matrix completion-based methods [2–4] stand out owing to their robustness and efficiency property since they avoid the image segmentation and features similarity calculation procedures.

Matrix completion-based methods usually operate on the tag matrix $\mathbf{O} \in \mathbb{R}^{N_{im} \times N_{tg}}$, where each row corresponds to one image, each column corresponds to one tag, and N_{im} and N_{tg} denote the number of images and tags respectively. $o_{ij} = 1$ only if image i is annotated with tag j and 0 otherwise. Thus one can get a completed tag matrix $\hat{\mathbf{O}}$ by completing the matrix $\mathbf{O}$ [2,5,6]. During the matrix completion procedure, if we can observe that o_{ij} is nonzero (zero) but $\hat{o}_{ij}$ becomes zero (nonzero), we say that the algorithm removes (adds) tag j from (to) image i.

L. Cheng et al. (Eds.): ISNN 2016, LNCS 9719, pp. 92–99, 2016.
DOI: 10.1007/978-3-319-40663-3_11

Semantically similar images usually have similar tags. However, the relationship between images and tags can hardly be characterized by linear models and traditional matrix completion-based methods can not take full advantage of side information such as user activity (e.g., like and reblog) and rich content (e.g., tags and images) [1]. In this work, we exploit the recently proposed BIMC [1] model, which combines the power of IMC [7] together with MC. BIMC have demonstrated its scalability and capability of exploiting side information in blog recommendation [1], where it effectively combines heterogeneous user and blog features from multiple sources for more accurate recommendations. To make the most of the side information, we model the visual-tag correlation and semantic-tag correlation properties in the BIMC model. Word vectors and CNN features are further utilized for extracting the tag and the visual features, respectively. These features have high level semantic meanings and can narrow the semantic gap effectively.

2 Boosted Inductive Matrix Completion

This section introduces the original BIMC model. Sections 2.1, 2.2 and 2.3 introduces the formulation for standard MC model, the IMC model and the BIMC model briefly.

2.1 Standard Matrix Completion

Low rank matrix completion (MC) recover the underlying low rank matrix by using the observed entries of $\mathbf{O}$, which is typically formulated as follows:

$$\min_{\mathbf{P},\mathbf{Q}} \|\mathbf{U} \odot (\mathbf{O} - (\mathbf{P}\mathbf{Q}^\top))\|_F^2 + \lambda(\text{rank}(\mathbf{P}\mathbf{Q}^\top)) \tag{1}$$

where $\mathbf{P} \in \mathbb{R}^{N_{im}}$ and $\mathbf{Q} \in \mathbb{R}^{N_{tg}}$ with r being the dimension of the latent feature space; $\mathbf{U}$ is the 0/1 binary mask with the same size as $\mathbf{O}$. The entry value 0 means that the corresponding entry in $\mathbf{O}$ is not observed, and 1 otherwise. The operator $\odot$ is the Hadamard entry-wise product. λ is a regularization parameter. The low-rank constraint on $\mathbf{P}\mathbf{Q}^\top$ is NP-hard to solve. The standard relaxation of the rank constraint is the trace norm, which is equivalent to minimizing $\frac{1}{2}(\|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2)$ [7]:

$$\min_{\mathbf{P},\mathbf{Q}} \|\mathbf{U} \odot (\mathbf{O} - (\mathbf{P}\mathbf{Q}^\top))\|_F^2 + \frac{\lambda}{2}(\|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2) \tag{2}$$

Note that MC only utilizes the observed entries of $\mathbf{O}$.

2.2 Inductive Matrix Completion

Standard MC methods is restricted to their transductive setting, thus cannot predict tags for new images. Further more, MC suffers performance with extreme

sparsity in the data [1]. IMC is proposed to alleviate data sparsity issues as well as enable predictions for new images and tags by incorporating side information.

Let $\mathbf{v}_i \in \mathbb{R}^{f_{im}}$ denote the feature vector of image i, and $\mathbf{t}_j \in \mathbb{R}^{f_{tg}}$ denote the feature vector of tag j. Let $\mathbf{V} \in \mathbb{R}^{N_{im} \times f_{im}}$ denote the feature matrix of N_{im} images, where the i-th row is the image feature vector $\mathbf{v}_i$, and $\mathbf{T} \in \mathbb{R}^{N_{tg} \times f_{tg}}$ denote the feature matrix of N_{tg} tags, where the i-th row is the tag feature $\mathbf{t}_i$.

IMC assume that the tag matrix is generated by applying feature vectors associated with its row as well as column entities to a underlying low-rank matrix $\mathbf{M} = \mathbf{W}\mathbf{H}^\top$, where $\mathbf{W} \in \mathbb{R}^{f_{im} \times r}, \mathbf{H} \in \mathbb{R}^{r \times f_{tg}}$ are of rank $r \ll N_{im}, N_{tg}$:

$$\min_{\mathbf{W},\mathbf{H}} \quad \text{loss}(\mathbf{O}_{\text{i,j}}, (\mathbf{V}\mathbf{W}\mathbf{H}^\top\mathbf{T}^\top)_{\text{i,j}}) + \frac{\lambda}{2}(\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2) \tag{3}$$

A common choice for the loss function is the squared loss:

$$\min_{\mathbf{W},\mathbf{H}} \|\mathbf{U} \odot (\mathbf{O} - \mathbf{V}\mathbf{W}\mathbf{H}^\top\mathbf{T}^\top)\|_F^2 + \frac{\lambda}{2}(\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2) \tag{4}$$

2.3 Boosted Inductive Matrix Completion

IMC is too rigid as it heavily depends on the image feature matrix $\mathbf{V}$ and tag feature matrix $\mathbf{T}$. BIMC tackle the problem by combine both standard MC and IMC, and thereby better utilize the power of both. BIMC combine the power of MC to reduce the noise level in the input data as well as the advantage of IMC to incorporate side information of users and items [1]. BIMC models $\mathbf{O}_{i,j}$ as

$$\mathbf{O}_{i,j} = (\mathbf{P}\mathbf{Q}^\top)_{i,j} + \alpha \mathbf{v_i}^\top \mathbf{M} \mathbf{t_j} \tag{5}$$

where the parameter α adjusts the contribution of features in the final prediction.

BIMC first learn the latent factor matrices $\mathbf{P}$ and $\mathbf{Q}$ of the MC model as in (2). The resulting approximation error or residual matrix $\mathbf{R} = \mathbf{O} - \mathbf{P}\mathbf{Q}^\top$ can then be modeled with IMC as:

$$\mathbf{R}_{i,j} = \mathbf{O}_{i,j} - (\mathbf{P}\mathbf{Q}^\top)_{i,j} = \mathbf{v_i}^\top \mathbf{M} \mathbf{t_j} \tag{6}$$

Thus, choosing the squared loss, the object function of BIMC is

$$\min_{\mathbf{W},\mathbf{H}} \quad \|\mathbf{U} \odot (\mathbf{O} - \mathbf{P}\mathbf{Q}^\top - \mathbf{V}\mathbf{W}\mathbf{H}^\top\mathbf{T}^\top)\|_F^2 + \frac{\lambda}{2}(\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2) \tag{7}$$

We will introduce the incorporation of visual-tag correlation and semantic-tag correlation in Sect. 3.

3 Incorporating Visual-Tag Correlation and Semantic-Tag Correlation

To make the most of side information, we incorporate the visual-tag correlation and semantic-tag correlation in the BIMC model.

Let the ith row of the residual matrix $\mathbf{R}$ as $\mathbf{R}_i$, corresponding to the residual tag vector of image i. Thus we can measure the correlation between image i and image j in two ways: (1) similarity between image features $\mathbf{v}_i$ and $\mathbf{v}_j$, (2) similarity between residual tag vectors $\mathbf{R}_i$ and $\mathbf{R}_j$. Since semantically similar images usually have similar tags, these two kinds of similarities should be correlated.

Similarly, since each column of the the residual matrix $\mathbf{R}$ represents the feature of a tag, we can measure the correlation between tag i and tag j in two ways: (1) similarity between their corresponding word vectors $\mathbf{t}_i$ and $\mathbf{t}_j$, (2) similarity between $\mathbf{R}^i$ and $\mathbf{R}^j$. These two kinds of similarities should be correlated, too.

We define $g_{ij} = \cos(\mathbf{v}_i, \mathbf{v}_j)$ and $h_{ij} = \cos(\mathbf{t}_i, \mathbf{t}_j)$ to measures the similarity between $\mathbf{v}_i, \mathbf{v}_j$ and $\mathbf{t}_i, \mathbf{t}_j$, respectively. Similar to [8], we model the two kinds of correlation using Graph Laplacian technique [9]:

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{H}} \mathrm{Tr}(\mathbf{R}\mathbf{L}_\mathbf{v}\mathbf{R}^\top + \mathbf{R}^\top\mathbf{L}_\mathbf{s}\mathbf{R}) = \\ &\min_{\mathbf{W},\mathbf{H}} [\mathrm{Tr}(\mathbf{V}\mathbf{W}\mathbf{H}^\top\mathbf{T}^\top\mathbf{L}_\mathbf{v}\mathbf{T}\mathbf{H}\mathbf{W}^\top\mathbf{V}^\top) + \mathrm{Tr}(\mathbf{T}\mathbf{H}\mathbf{W}^\top\mathbf{V}^\top\mathbf{L}_\mathbf{s}\mathbf{V}\mathbf{W}\mathbf{H}^\top\mathbf{T}^\top)] \end{aligned} \tag{8}$$

where $\mathbf{L}_\mathbf{v} = \mathrm{diag}(\mathbf{G}\mathbf{1}) - \mathbf{G}$ is the Graph Laplacian matrix of visual similarity matrix $\mathbf{G}$, and $\mathbf{L}_\mathbf{s} = \mathrm{diag}(\mathbf{H}\mathbf{1}) - \mathbf{H}$ is the Graph Laplacian matrix of semantic similarity matrix $\mathbf{H}$.

Thus we can incorporate the two kinds of correlation into IMC:

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{H}} \quad \|\mathbf{U} \odot (\mathbf{O} - \mathbf{P}\mathbf{Q}^\top - \mathbf{V}\mathbf{W}\mathbf{H}^\top\mathbf{T}^\top)\|_F^2 + \frac{\lambda_1}{2}(\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2) + \\ &\lambda_2[\mathrm{Tr}(\mathbf{V}\mathbf{W}\mathbf{H}^\top\mathbf{T}^\top\mathbf{L}_\mathbf{v}\mathbf{T}\mathbf{H}\mathbf{W}^\top\mathbf{V}^\top) + \mathrm{Tr}(\mathbf{T}\mathbf{H}\mathbf{W}^\top\mathbf{V}^\top\mathbf{L}_\mathbf{s}\mathbf{V}\mathbf{W}\mathbf{H}^\top\mathbf{T}^\top)] \end{aligned} \tag{9}$$

We set a same weight parameter λ_2 for both visual-tag correlation and semantic-tag correlation for optimization efficiency.

4 Optimization

The objective function is non-convex. We utilize the same MC and IMC solvers adopted in [1] to solve our formulation. First we solve MC subproblem using [10], then we use LELM [11] method, which naturally fits for the large-scale multi-label learning with missing labels task, to solve the improved IMC subproblem. The solver uses alternating minimization (fix $\mathbf{W}$ and solve for $\mathbf{H}$ and vice versa) to optimize the function. When $\mathbf{W}$ or $\mathbf{H}$ is fixed, the resulting problem in one variable ($\mathbf{H}$ or $\mathbf{W}$) is solved using the Conjugate Gradient iterative procedure. For example, fixing $\mathbf{H}$, the gradient of the above objective in matrix form is given as:

$$\begin{aligned} &2\mathbf{U} \odot \mathbf{V}^\top[\mathbf{U} \odot (\mathbf{O} - \mathbf{P}\mathbf{Q}^\top - \mathbf{V}\mathbf{W}\mathbf{H}^\top\mathbf{T}^\top)]\mathbf{T}\mathbf{H} + \lambda_1\mathbf{W} + \\ &2\lambda_2(\mathbf{V}^\top\mathbf{V}\mathbf{W}\mathbf{H}\mathbf{T}^\top\mathbf{L}_\mathbf{v}\mathbf{T}\mathbf{H} + \mathbf{V}^\top\mathbf{L}_\mathbf{S}\mathbf{V}\mathbf{W}\mathbf{H}\mathbf{T}^\top\mathbf{T}\mathbf{H}) \end{aligned} \tag{10}$$

5 CNN Features and Semantic Vectors

We utilize DeCAF_6 [12] to extract visual features, which have high level semantic meanings thus are more representative than low level visual features. And we adopt pre-trained word2vec [13] to calculate the word vectors for each tag, which could keep their semantic meanings precisely.

6 Experimental Evaluation

The proposed model is denoted as BITMC (Boosted Inductive Tag Matrix Completion). We follow the same experimental settings in [8] and evaluate BITMC on three benchmark datasets: Corel5K, Labelme [14] and MIRFlickr-25K [15].

6.1 Datasets and Experimental Setup

LabelMe dataset is collected through an online tagging project. MIRFlickr-25K is collected from Flickr. Compared to Corel5K and Labelme, tags in MIRFlickr-25K are much more noisy. Hence, a pre-processing procedure is performed. We match each tag with entries in a Wikipedia thesaurus and only retain the tags in accordance with Wikipedia. We extract tag vectors and visual features for all the datasets.

To study the tag completion performance, multiple models are employed as the baselines, including matrix completion-based models (LRES [5], TCMR [3], RKML [4], 4 Priors* [8] and ITMC), search-based models (JEC [16], TagProp [17] and TagRelevance [18]), mixture models (CMRM [19] and MBRM [20]) and CCA based model FastTag [21]. Note that we denote our model without the MC part as ITMC. 4 Priors is based on IMC, incorporating the same visual-tag correlation, semantic-tag correlation and inhomogeneous errors properties [8]. For the sake of fair comparison, we remove the inhomogeneous errors term to get the 4 Priors*.

We tuned λ_1, λ_2 using cross validation, and the parameters of adopted baselines are also carefully tuned on the validation set of the three datasets using the same strategies as in [6].

We adopt the same evaluation metrics used in [8]. All models are evaluated in terms of *average precision*@N (i.e. $AP@N$), *average recall*@N (i.e. $AR@N$) and *coverage*@N (i.e. $C@N$). In the top N completed tags, *precision*@N is to measure the ratio of correct tags in the top N competed tags and *recall*@N is to measure the ratio of missing ground-truth tags, both averaged over all test images. *Coverage*@N is to measure the ratio of test images with at least one correctly completed tag.

6.2 Evaluation and Observation

Tables 1, 2 and 3 show performance comparisons on the three datasets. Top 3 in each measure is shown in bold.

Table 1. Performance comparison on Corel5K

	Corel5K											
	N = 2			N = 3			N = 5			N = 10		
	AP	**AR**	**C**	**AP**	**AR**	**C**	**AP**	**AR**	**C**	**AP**	**AR**	**C**
BITMC	**0.58**	**0.43**	**0.52**	**0.49**	**0.48**	**0.64**	**0.44**	**0.56**	**0.65**	**0.38**	**0.62**	**0.89**
ITMC	0.56	**0.42**	0.49	0.46	**0.49**	0.57	0.41	**0.55**	0.63	0.35	**0.64**	**0.88**
4 Priors* [8]	**0.58**	**0.41**	**0.50**	**0.48**	**0.49**	**0.62**	**0.42**	**0.58**	**0.65**	**0.37**	**0.62**	**0.91**
LRES [5]	**0.58**	0.39	0.47	**0.48**	**0.48**	0.57	0.41	0.53	0.62	**0.37**	**0.62**	0.85
TCMR [3]	0.57	0.39	0.49	**0.48**	0.47	**0.58**	**0.44**	**0.55**	**0.66**	**0.38**	0.61	**0.88**
RKML [4]	0.29	0.21	0.24	0.25	0.24	0.29	0.23	0.25	0.34	0.19	0.29	0.67
JEC [16]	0.36	0.34	0.39	0.31	0.40	0.47	0.27	0.32	0.59	0.20	0.33	0.76
TagProp [17]	0.46	0.40	**0.50**	0.38	**0.48**	0.57	0.33	0.51	0.63	0.26	0.54	0.86
TagRel [18]	0.43	**0.41**	0.48	0.37	0.47	0.57	0.31	0.50	0.60	0.26	0.53	0.90
CMRM [19]	0.29	0.20	0.23	0.24	0.24	0.27	0.21	0.25	0.35	0.16	0.27	0.63
MBRM [20]	0.35	0.29	0.35	0.28	0.34	0.42	0.24	0.24	0.39	0.17	0.28	0.70
FastTag [21]	0.54	0.31	0.45	0.46	0.44	0.51	0.40	0.52	0.63	0.36	**0.62**	0.82

Table 2. Performance comparison on Labelme

	Labelme											
	N = 2			N = 3			N = 5			N = 10		
	AP	**AR**	**C**	**AP**	**AR**	**C**	**AP**	**AR**	**C**	**AP**	**AR**	**C**
BITMC	**0.47**	**0.34**	**0.41**	**0.48**	**0.37**	**0.51**	**0.45**	**0.47**	**0.60**	**0.34**	**0.60**	**0.77**
ITMC	0.43	**0.35**	0.39	0.43	0.35	0.48	0.36	**0.46**	0.58	**0.32**	**0.57**	**0.77**
4 Priors* [8]	**0.50**	**0.35**	**0.42**	**0.47**	**0.39**	**0.52**	**0.42**	**0.47**	**0.61**	**0.32**	**0.60**	0.74
LRES [5]	0.42	0.32	0.39	0.40	0.36	0.50	0.35	0.45	0.55	0.27	0.56	0.69
TCMR [3]	**0.44**	0.32	**0.42**	0.41	0.36	**0.51**	**0.37**	0.45	**0.60**	0.29	0.55	**0.75**
RKML [4]	0.21	0.14	0.20	0.20	0.16	0.21	0.19	0.20	0.23	0.14	0.22	0.28
JEC [16]	0.33	0.29	0.31	0.30	0.32	0.37	0.27	0.38	0.45	0.20	0.48	0.58
TagProp [17]	0.39	0.31	0.36	0.35	**0.37**	0.45	0.33	0.45	0.52	0.25	0.56	0.64
TagRel [18]	0.43	0.32	0.36	0.37	0.35	0.44	0.34	0.45	0.51	0.27	0.55	0.62
CMRM [19]	0.20	0.14	0.18	0.18	0.15	0.20	0.18	0.19	0.25	0.12	0.22	0.29
MBRM [20]	0.23	0.14	0.18	0.21	0.16	0.21	0.18	0.20	0.25	0.12	0.27	0.37
FastTag [21]	0.43	**0.34**	0.40	**0.48**	0.36	0.44	**0.37**	0.44	0.53	0.28	**0.57**	0.70

We can observe that methods achieve better performance on Corel5K and Labelme than MIRFlickr-25K, since tags in MIRFlickr-25K are much more noisy. Matrix completion-based semi-supervised methods, such as BITMC, LRES, TCMR, ITMC and 4 Priors* usually achieve the best performances owing to the advantage of exploiting both labeled (few) and large number of unlabeled information. In all cases, BITMC, ITMC and 4 Priors* achieve satisfactory performances. BITMC combines the power of ITMC and standard MC, which is

Table 3. Performance comparison on MIRFlickr-25K

	MIRFlickr-25K											
	N = 2			N = 3			N = 5			N = 10		
	AP	**AR**	**C**	**AP**	**AR**	**C**	**AP**	**AR**	**C**	**AP**	**AR**	**C**
BITMC	**0.50**	**0.36**	**0.44**	**0.48**	**0.43**	**0.54**	**0.38**	**0.44**	**0.60**	**0.32**	**0.61**	**0.81**
ITMC	**0.45**	**0.36**	**0.43**	**0.44**	0.41	**0.53**	**0.37**	**0.44**	0.56	**0.29**	**0.58**	**0.78**
4 Priors* [8]	**0.52**	**0.35**	0.41	**0.47**	0.40	0.50	**0.38**	**0.43**	0.57	**0.29**	**0.56**	0.74
LRES [5]	0.43	**0.35**	0.40	0.40	0.39	**0.53**	0.32	0.40	0.57	0.26	0.45	0.73
TCMR [3]	**0.45**	**0.35**	**0.44**	0.43	0.38	**0.54**	0.35	0.41	**0.60**	0.28	0.48	0.77
RKML [4]	0.21	0.15	0.15	0.23	0.22	0.25	0.13	0.23	0.31	0.13	0.22	0.55
JEC [16]	0.33	0.30	0.32	0.31	0.38	0.45	0.25	0.34	0.55	0.19	0.35	0.66
TagProp [17]	0.39	0.35	0.39	0.36	0.42	0.51	0.28	0.37	**0.59**	0.20	0.41	0.73
TagRel [18]	0.42	0.34	0.37	0.37	**0.43**	0.52	0.30	0.37	0.57	0.20	0.40	**0.78**
CMRM [19]	0.20	0.15	0.16	0.18	0.21	0.24	0.13	0.18	0.30	0.11	0.20	0.50
MBRM [20]	0.22	0.16	0.18	0.17	0.30	0.35	0.13	0.18	0.33	0.10	0.22	0.55
FastTag [21]	0.43	0.35	0.38	0.39	**0.43**	0.51	0.30	0.41	0.57	0.27	0.42	0.75

verified on the performance comparisons between BITMC and ITMC. Note that as the dataset become more noisy, their difference becomes larger. The reason for this phenomenon is that as the data becomes noisier, the benefit of side information (IMC) becomes relatively small comparing to the benefit of low-rankness (MC). Note that BITMC is more efficient than 4 Priors* because BIMC do not have to explicitly form $\mathbf{O} - \mathbf{P}\mathbf{Q}^\top$ [1] and the optimization procedure has closed-form solutions.

7 Conclusion

We have improved the powerful BIMC model and proposed an effective model BITMC for tag completion, which takes low-rankness, visual-tag correlation, semantic-tag correlation into consideration. We utilize word vectors to calculate semantic-tag correlation and CNN features to measure tag-visual correlation. BITMC outperforms several state-of-the-art methods on benchmark datasets.

References

1. Shin, D., Cetintas, S., Lee, K., Dhillon, I.: Tumblr blog recommendation with boosted inductive matrix completion. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. ACM (2015)
2. Goldberg, A., Recht, B., Xu, J., Nowak, R., Zhu, X.: Transduction with matrix completion: three birds with one stone. In: Advances in Neural Information Processing Systems (2010)
3. Feng, Z., Feng, S., Jin, R., Jain, A.K.: Image tag completion by noisy matrix recovery. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part VII. LNCS, vol. 8695, pp. 424–438. Springer, Heidelberg (2014)

4. Feng, Z., Jin, R., Jain, A.: Large-scale image annotation by efficient and robust kernel metric learning. In: Proceedings of the IEEE International Conference on Computer Vision (2013)
5. Zhu, G., Yan, S., Ma, Y.: Image tag refinement towards low-rank, content-tag prior and error sparsity. In: Proceedings of the International Conference on Multimedia. ACM (2010)
6. Wu, L., Jin, R., Jain, A.: Tag completion for image retrieval. IEEE Trans. Pattern Anal. Mach. Intell. **35**(3), 716–727 (2013)
7. Jain, P., Dhillon, I.: Provable inductive matrix completion (2013). arXiv preprint arXiv:1306.0626
8. Hou, Y.: Image annotation incorporating low-rankness, tag and visual correlation and inhomogeneous errors. In: Jiang, J., et al. (eds.) ISVC 2015. LNCS, vol. 9474, pp. 71–81. Springer, Heidelberg (2015). doi:10.1007/978-3-319-27857-5_7
9. Chung, F.: Spectral Graph Theory. American Mathematical Society, Providence (1997)
10. Jain, P., Netrapalli, P., Sanghavi, S.: Low-rank matrix completion using alternating minimization. In: Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing. ACM (2013)
11. Yu, H., Jain, P., Kar, P., Dhillon, I.: Large-scale multi-label learning with missing labels. In: Proceedings of The 31st International Conference on Machine Learning (2014)
12. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: DeCAF: a deep convolutional activation feature for generic visual recognition (2013). arXiv preprint arXiv:1310.1531
13. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space (2013). arXiv preprint arXiv:1301.3781
14. Russell, B., Torralba, A., Murphy, K., Freeman, W.: Labelme: a database and web-based tool for image annotation. Int. J. Comput. Vis. **77**(1), 157–173 (2008)
15. Huiskes, M., Lew, M.: The MIR flickr retrieval evaluation. In: Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval. ACM (2008)
16. Makadia, A., Pavlovic, V., Kumar, S.: A new baseline for image annotation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 316–329. Springer, Heidelberg (2008)
17. Guillaumin, M., Mensink, T., Verbeek, J., Schmid, C.: TagProp: discriminative metric learning in nearest neighbor models for image auto-annotation. In: Proceedings of the IEEE 12th International Conference on Computer Vision, pp. 309–316 (2009)
18. Li, X., Snoek, C., Worring, M.: Learning social tag relevance by neighbor voting. IEEE Trans. Multimedia **11**(7), 1310–1322 (2009)
19. Jeon, J., Lavrenko, V., Manmatha, R.: Automatic image annotation and retrieval using cross-media relevance models. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval. ACM (2003)
20. Feng, S., Manmatha, R., Lavrenko, V.: Multiple Bernoulli relevance models for image and video annotation In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2004)
21. Chen, M., Zheng, A., Weinberger, K.: Fast image tagging. In: Proceedings of the 30th International Conference on Machine Learning (2013)