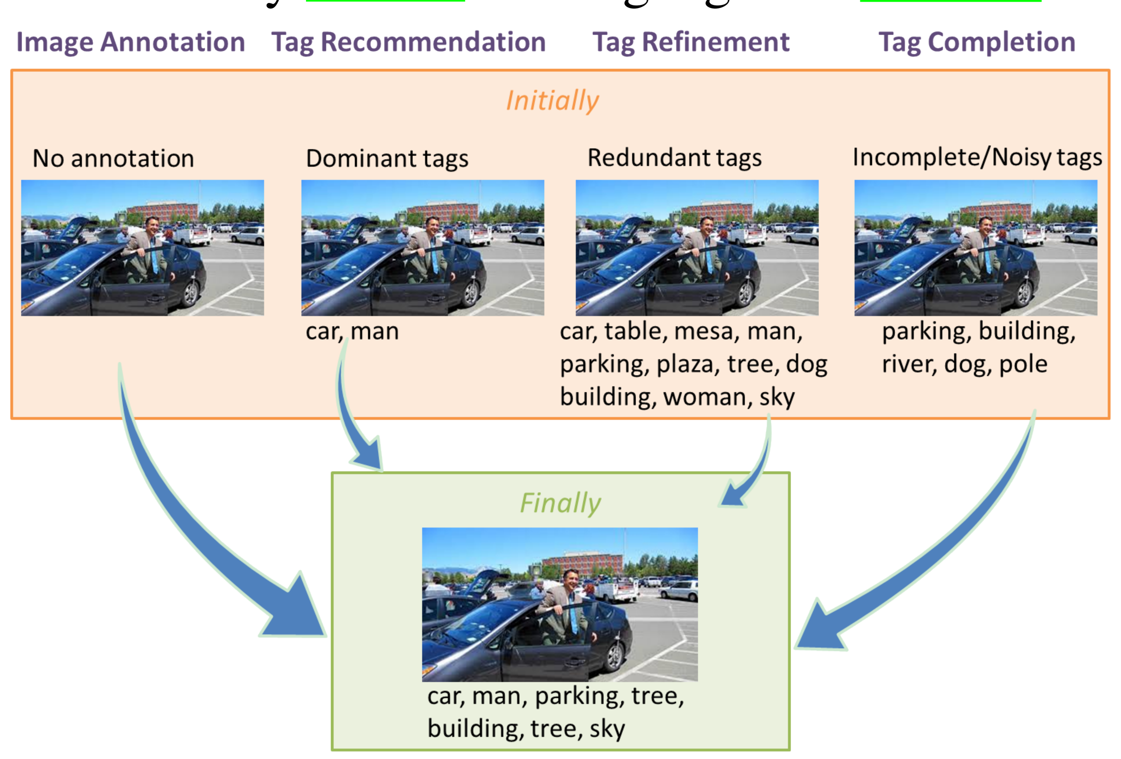
Image Tag Completion by Noisy Matrix Recovery

Zheyun Feng*, Songhe Feng‡, Rong Jin*, Anil K. Jain*

*Michigan State University, ‡Beijing Jiaotong University

Background and Motivation

Simultaneously **enrich** missing tags and **remove** noisy tags for images.



Motivation:

✦ Large amount of images with incomplete and inaccurate tags;

✦ Popularity of tag based tasks, eg., tag-based image retrieval.

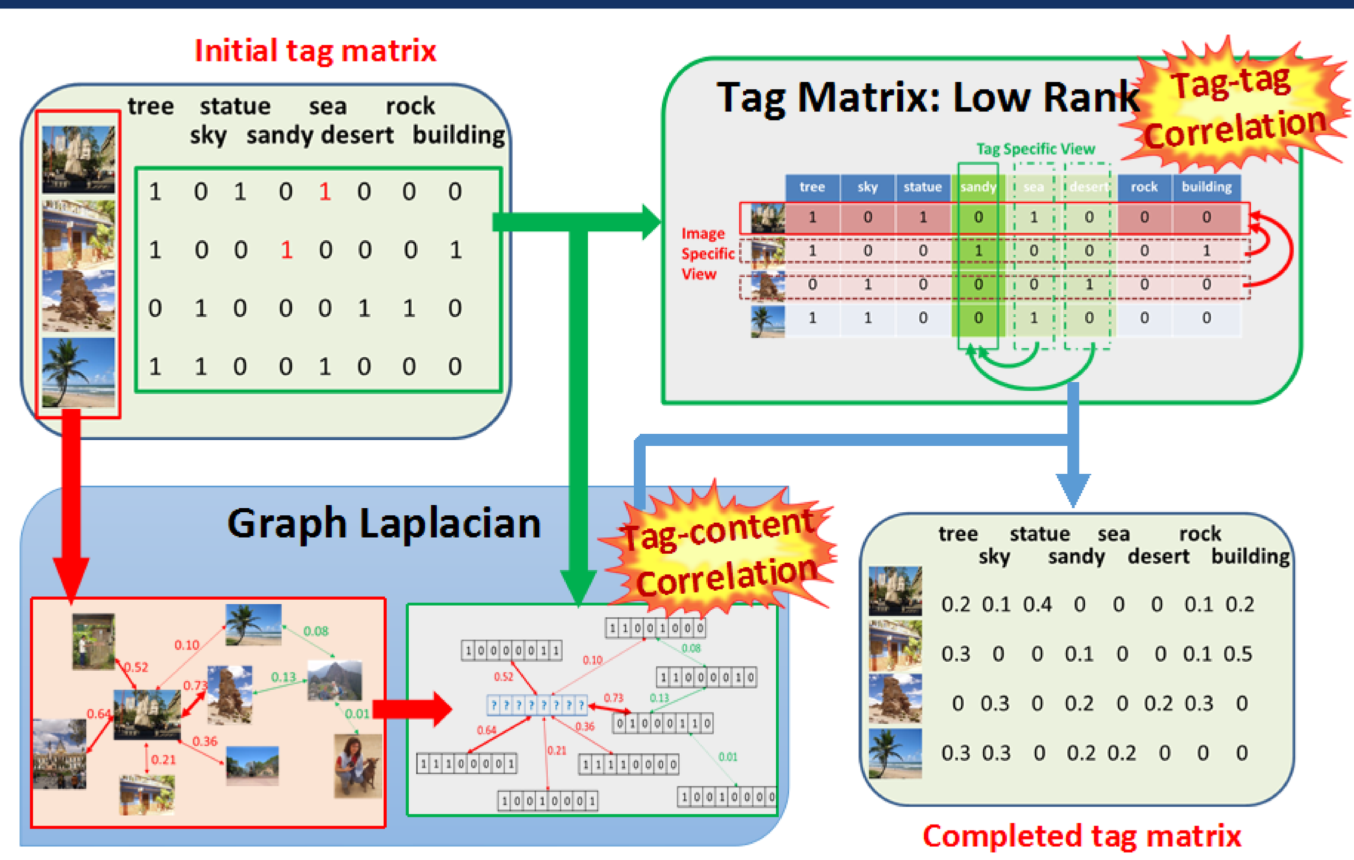
Limitations of existing image tagging algorithms:

✧ Dealing with only one of the two problems;

✧ Large amount of training images with complete and accurate tags;

✧ No principled approach of capturing the correlation among tags.

Proposed TCMR Algorithm



☞ Convex optimization ➟ computationally efficient;

☞ Low rank enforcement ➟ key assumption in topic model;

☞ Graph Laplacian exploration ➟ consistent between tags and visual cues;

☞ Provide theoretical guarantee for image tag completion for the first time.

Two Assumptions

✌ Idea of Language Model: Observed tags of each image are drawn independently from a multinomial distribution.

- ☛ Number of observed tags (m_*) is limited;
- ☛ Number of parameters to be estimated is significantly larger than m_*.

✌ Low Rank Matrix Recovery: Tags of any image are sampled from a mixture of *a small number* of multinomial distribution.

➟ Recovered tag matrix has to be of *low rank*.

Notation

- m, m_*: the number of unique tags or assigned tags for each image;
- $\mathcal{D} = \{\mathbf{d}_1, \cdots, \mathbf{d}_n\}$: tagged image set, where $\mathbf{d}_i$ is the i-th tag vector;
- $P = (\mathbf{p}_1, \cdots, \mathbf{p}_n)$: the multinomial distributions for all images;
- $\mathbf{p}_i$: the multinomial distribution to generate tags in $\mathbf{d}_i$;
- $|Q|_{tr}$, $|Q|_1$: the nuclear (trace) norm and ℓ_1 norm of matrix.

Tag Completion by Noisy Matrix Recovery (TCMR)

Recover the multinomial probability P by combining the maximum likelihood estimation and low rank matrix recovery theory

$$\min_{Q\in\Delta} \quad \mathcal{L}(Q) := -\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{d_{i,j}}{m_*}\log Q_{i,j} + \varepsilon|Q|_{tr}.$$

where $\Delta = \left\{Q \in (0,1)^{m\times n} : Q_{*,i}^\top \mathbf{1} = 1, i \in [1,n]\right\}$.

☞ Left term: ensures consistency between optimal $\hat{Q}$ and observed tags;

☞ Right term: enforces tag matrix to be low rank.

✔ Entries sampled from unknown multinomial distributions ➟ likelihood;

✗ Entries sampled uniformly at random from a given matrix ➟ square loss.

Theoretical Guarantee of RKML

Theorem. *Let r be the rank of matrix P, N be the total number of observed tags, and $\hat{Q}$ be the optimal P. Assume $N \ge \Omega(n\log(n+m))$, and denote by μ_- and μ_+ the lower and upper bounds for the probabilities in P. Then we have, with a high probability*

$$\frac{1}{n}|\hat{Q} - P|_1 \le O\left(\frac{rn\theta^2\log(n+m)}{N}\right), \text{ where } \theta^2 := \frac{\mu_+|P\vec{1}|_\infty}{n\mu_-^2} \le \frac{\mu_+^2}{\mu_-^2}.$$

☞ Recovery error: $O(rn\log(n+m)/N)$;

☞ Tag matrix can be accurately recovered when $N \ge \Omega(rn\log(n+m))$.

Incorporation with Visual Features and Irrelevant Tags

Optimization problem becomes:

$$\min_{Q\in\Delta} -\sum_{i,j=1}^{n,m}\left\{\frac{d_{i,j}}{m_*}\log Q_{i,j} + \frac{1-d_{i,j}}{m-m_*}\log(1-Q_{i,j})\right\} + \frac{\alpha}{n}Tr(Q^\top LQ) + \beta|Q|_{tr}.$$

☛ $X = (\mathbf{x}_1, \cdots, \mathbf{x}_n)^\top$: visual features of n images, where $\mathbf{x}_i \in \mathbb{R}^d$;

☛ $W = [w_{i,j}]_{n\times n}$: $w_{i,j} = \exp(-d(\mathbf{x}_i, \mathbf{x}_j)^2/\sigma^2)$ is the pairwise similarity;

☛ $L = \mathrm{diag}(W^\top\mathbf{1}) - W$: the graph Laplacian;

☛ $Tr(Q^\top LQ) = \sum_{i,j=1}^{n} W_{i,j}|Q_{*,i} - Q_{*,j}|^2$: tag-visual content correlation.

Efficient Solution

Re-write the objective function as $\mathcal{L}(Q) = f(Q) + \varepsilon|Q|_{tr}$, and

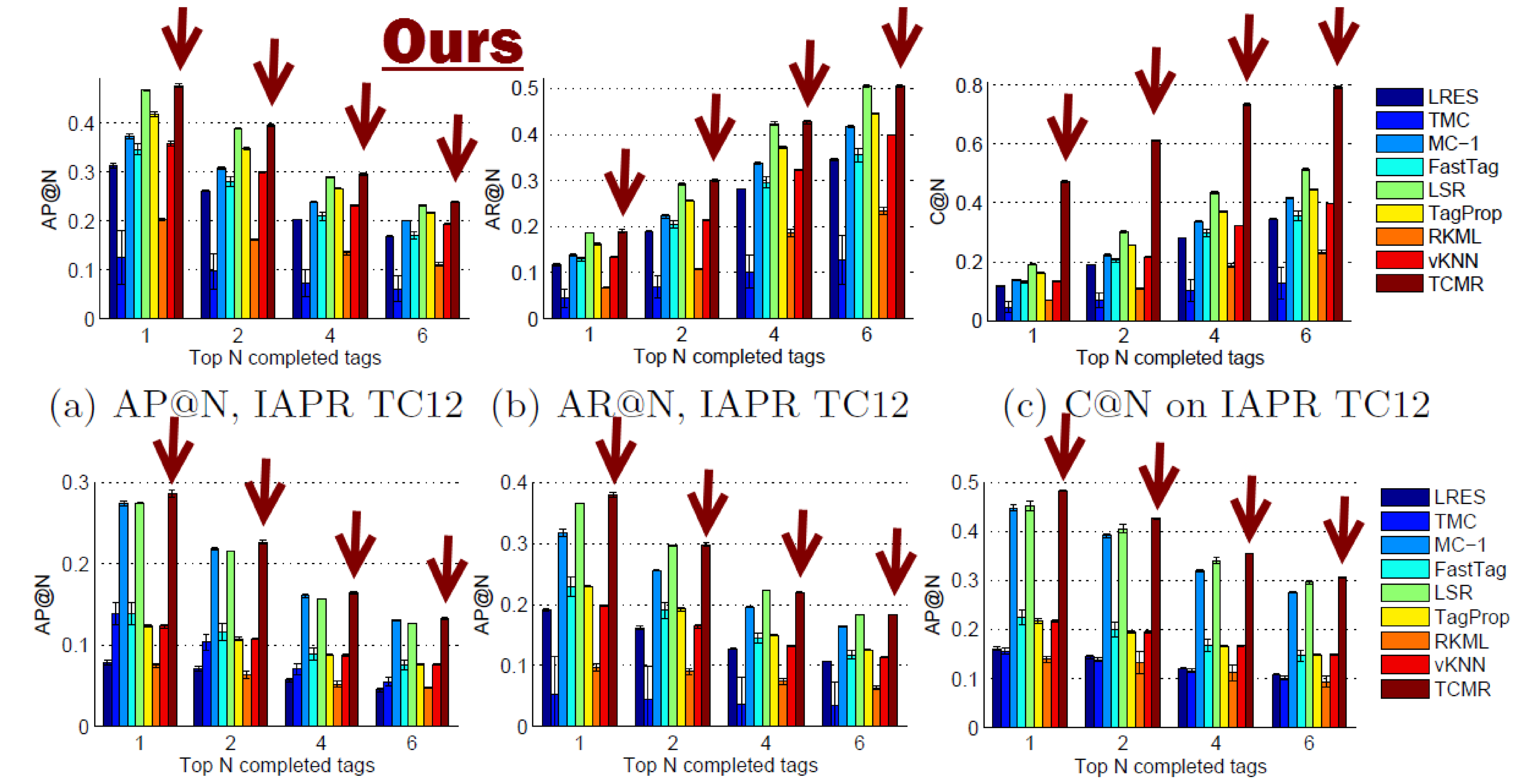
$$Q_k = \arg\min_Q \; P_{t_k}(Q, Q_{k-1}) = \frac{1}{2}|Q - (Q_{k-1} - \frac{1}{t_k}\nabla f(Q_{k-1}))|_F^2 + \frac{\varepsilon}{t_k}|Q|_{tr}.$$

where t_k is the step size for the kth iteration.

Datasets

	No. of imgs	No. of tags	Mean, min, max tags per img			Mean, min, max imgs per tag			N Observed tags
Mir Flickr	5,231	372	5.82	4	43	82	10	781	3
ESP Game	10,450	265	6.41	5	15	253	16	3,439	4
IAPR TC12	12,985	291	7.07	5	23	316	14	4,752	4
NUS-WIDE	20,968	420	10.4	9	15	520	78	5,058	4

Comparison with State-of-the-art Baselines



(a) AP@N, IAPR TC12 (b) AR@N, IAPR TC12 (c) C@N on IAPR TC12

(d) AP@N on Mir Flickr (e) AP@N on ESP Game (f) AP@N on NUSWIDE

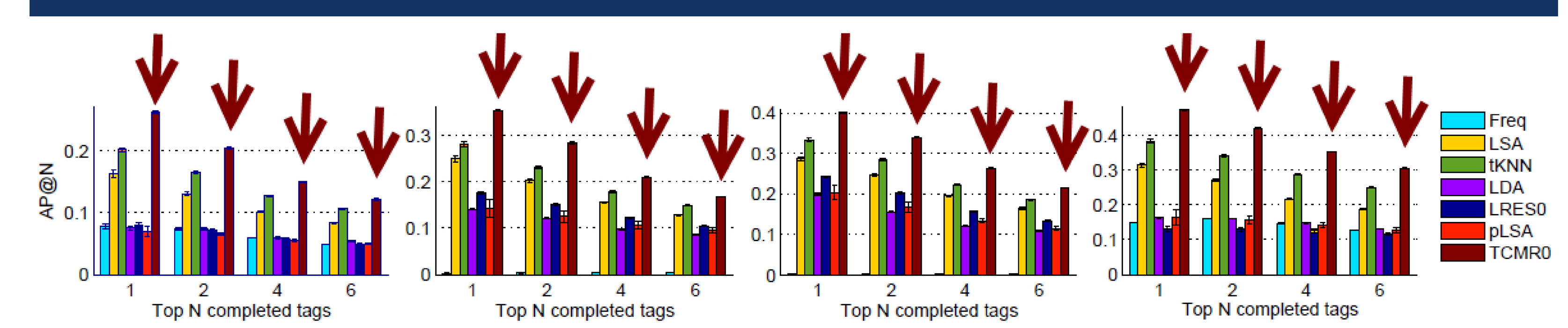
Fig. 1. Tag completion performance of TCMR and state-of-the-art baselines.

Table 1. Running time (seconds) for tag completion baselines. All algorithms are run in Matlab on an AMD 4-core @2.7GHz and 64GB RAM machine.

Best baseline — Ours

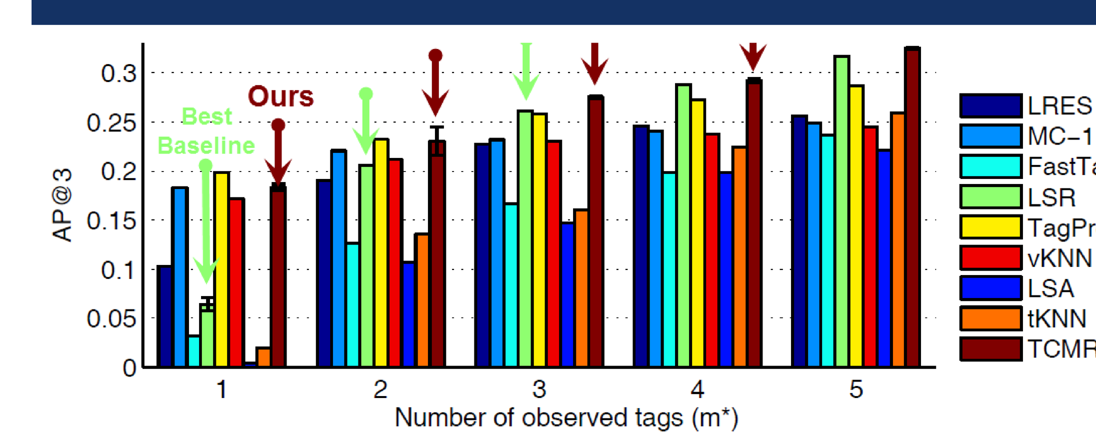
	LRES	TMC	MC-1	FastTag	LSR	TagProp	RKML	vKNN	TCMR
MirFlickr	5.6e2	4.7e3	8.6e2	1.4e3	6.2e3	2.5e2	3.0e2	2.1e2	1.3e3
ESP Game	3.4e2	5.8e3	1.0e3	8.6e2	1.3e4	6.7e2	1.3e3	4.3e2	5.9e3
IAPR TC12	5.2e2	1.2e4	1.7e3	1.6e3	1.6e4	1.1e3	1.5e3	1.0e3	9.4e3
NUS-WIDE	6.8e3	2.9e4	1.8e3	2.6e3	2.8e4	1.5e3	3.8e3	1.2e3	1.9e4

Comparison without Considering Visual Features



(a) Mir Flickr (b) ESP Game (c) IAPR TC12 (d) NUSWIDE

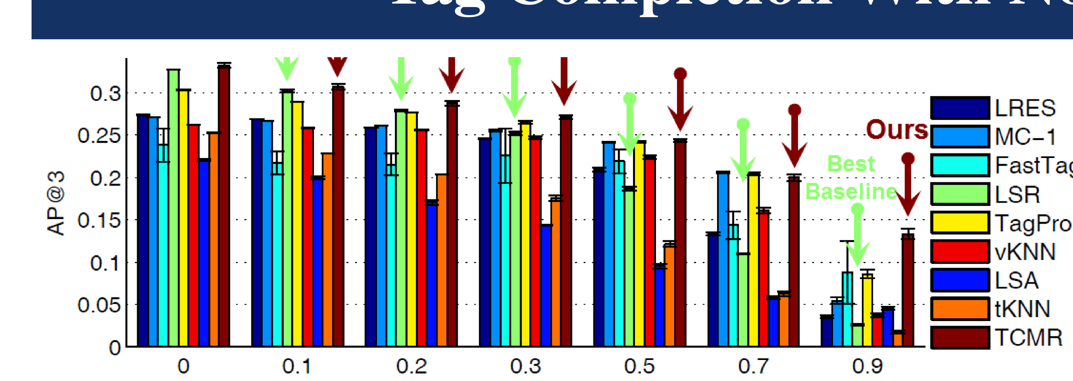
With Varied Number of Observed Tags



(a) $AP@3$ on IAPR TC12

1. As number of observed tags increases, tag completion performance increases;
2. When m*≥3, TCMR and LSR are super;
3. When m*≤2, TCMR outperforms LSR;
4. TCMR is better considering the rage of m*.

Tag Completion With Noisy Tags



(a) IAPR TC12 $AP@3$

1. As number of noisy tags increases, tag completion performance degrades.
2. LSR is more sensitive to noise, and performs bad when the noise is large.
3. TCMR is better, especially when recovering from severely noisy tags.