Image Tag Completion by Noisy Matrix Recovery

Supplementary Document

Zheyun Feng*, Songhe Feng‡, Rong Jin*, Anil K. Jain*
*{fengzhey, rongjin, jain}@cse.msu.edu, ‡shfeng@bjtu.edu.cn

*Michigan State University, ‡Beijing Jiaotong University

Abstract. In this supplementary document, we present

- Detailed proofs of Lemma 1, Lemma 2, theorem 2, theorem 4 and theorem 5 in the main paper.
- Detailed statistics about the refined datasets.
- Supplementary experimental results, mainly in terms of $AR@N$ and $C@N$.

Note all the notations are the same as used in the main paper.

1 Detailed Proofs

1.1 Proof of Lemma 1

Proof. We have

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|P_{i,j}-Q_{i,j}|^2}{Q_{i,j}} \\
=&\sum_{i=1}^{n}\left(\sum_{j=1}^{m}\frac{|P_{i,j}-Q_{i,j}|^2}{Q_{i,j}}\right)\left(\sum_{i=1}^{j}Q_{i,j}\right) \\
\geq&\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{|P_{i,j}-Q_{i,j}|}{\sqrt{Q_{i,j}}}\sqrt{Q_{i,j}}=|P-Q|_1.
\end{aligned}$$

1.2 Proof of Lemma 2

Proof. To facilitate our analysis, we rewrite each $\mathbf{d}_i$ as

$$\mathbf{d}_i=\sum_{j=1}^{m_*}\mathbf{d}_i^j,$$

where $\mathbf{d}_i^j$ is the image tag vector corresponding to the j-th word sampling for the tag vector of the i-th image. To utilize Lemma 2, we define $Z_{i,j}$ as

$$Z_i=\left(\mathbf{d}_i^j-\mathbf{p}_i\right)\mathbf{e}_i^\top,$$

and therefore

$$M = \frac{1}{m_*} \sum_{i=1}^{n} \sum_{j=1}^{m_*} Z_{i,j}.$$

To bound U in Lemma 2, we have

$$|Z_{i,j}|_* \leq \left| \mathbf{d}_i^j - \mathbf{p}_i \right|_2 \leq |\mathbf{d}_i^j|_2 \leq 1.$$

To bound σ_Z, we compute

$$\left| \frac{1}{nm_*} \sum_{i=1}^{n} \sum_{j=1}^{m_*} \mathbb{E}\left[Z_{i,j} Z_{i,j}^\top \right] \right|_* = \left| \frac{1}{nm_*} \sum_{i=1}^{n} \sum_{j=1}^{m_*} \mathbb{E}\left[\left(\mathbf{d}_i^j - \mathbf{p}_i \right) \left(\mathbf{d}_i^j - \mathbf{p}_i \right)^\top \right] \right|_*$$
$$= \left| \frac{1}{nm_*} \sum_{i=1}^{n} \sum_{j=1}^{m_*} \mathbb{E}\left[\mathbf{d}_i^j (\mathbf{d}_i^j)^\top \right] - \mathbf{p}_i \mathbf{p}_i^\top \right|_* \leq \max_{1 \leq j \leq m} \frac{1}{n} \sum_{i=1}^{n} p_{i,j}(i - p_{i,j}^2) = \frac{|P\mathbf{1}|_\infty}{n}.$$

Similarly, we have

$$\left| \frac{1}{nm_*} \sum_{i=1}^{n} \sum_{j=1}^{m_*} \mathbb{E}\left[Z_i^\top Z_i \right] \right|_* = \left| \frac{1}{nm_*} \sum_{i=1}^{n} \sum_{j=1}^{m_*} \mathbb{E}\left[\left(\mathbf{d}_i^j - \mathbf{p}_i \right)^\top \left(\mathbf{d}_i^j - \mathbf{p}_i \right) \mathbf{e}_i \mathbf{e}_i^\top \right] \right|_*$$
$$= \left| \frac{1}{nm_*} \sum_{i=1}^{n} \sum_{j=1}^{m_*} \mathbb{E}\left[\left(\mathbf{d}_i^\top \mathbf{d}_i - \mathbf{p}_i^\top \mathbf{p}_i \right) \mathbf{e}_i \mathbf{e}_i^\top \right] \right|_* \leq \frac{1}{n}.$$

We complete the proof by plugging the bounds for U and σ_Z.

1.3 Proof of Theorem 2

Proof. We consider any solution $Q \in \Delta$. Since $\hat{Q}$ is the optimal solution to Eq. (1) in the main paper, we have $\langle \nabla \mathcal{L}(\hat{Q}), \hat{Q} - Q \rangle \leq 0$, i.e.

$$-\frac{1}{m_*} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{d_{i,j}}{\hat{Q}_{i,j}} \left(\hat{Q}_{i,j} - Q_{i,j} \right) + \varepsilon \langle \partial |\hat{Q}|_{tr}, \hat{Q} - Q \rangle \leq 0,$$

where $\partial |\hat{Q}|_{tr}$ is a subgradient of $|\hat{Q}|_{tr}$. Using the fact that

$$\langle \partial |\hat{Q}|_{tr} - \partial |Q|_{tr}, \hat{Q} - Q \rangle \geq 0,$$

we can replace $\langle \partial |\hat{Q}|_{tr}, \hat{Q} - Q \rangle$ with $\langle \partial |Q|_{tr}, \hat{Q} - Q \rangle$, which results in the following inequality

$$-\frac{1}{m_*} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{d_{i,j}}{\hat{Q}_{i,j}} \left(\hat{Q}_{i,j} - Q_{i,j} \right) + \varepsilon \langle \partial |Q|_{tr}, \hat{Q} - Q \rangle \leq 0.$$

Define $Z_{i,j} = (\hat{Q}_{i,j} - Q_{i,j})/\hat{Q}_{i,j}$. We have

$$-\frac{1}{m_*}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{d_{i,j}}{\hat{Q}_{i,j}}\left(\hat{Q}_{i,j} - Q_{i,j}\right) = -\frac{1}{m_*}\sum_{i=1}^{n}\langle \mathbf{d}_i \mathbf{e}_i^\top, Z\rangle = -\langle P, Z\rangle - \langle M, Z\rangle.$$

Thus the bound in Eq. (8) in the main paper is modified as

$$-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{P_{i,j}}{\hat{Q}_{i,j}}\left(\hat{Q}_{i,j} - Q_{i,j}\right) + \varepsilon\langle \partial|Q|_{tr}, \hat{Q} - Q\rangle \leq \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{M_{i,j}}{\hat{Q}_{i,j}}\left(\hat{Q}_{i,j} - Q_{i,j}\right).$$

Since

$$-\sum_{j=1}^{m}\frac{P_{i,j}}{\hat{Q}_{i,j}}\left(\hat{Q}_{i,j} - Q_{i,j}\right) = -\sum_{j=1}^{m}\frac{1}{\hat{Q}_{i,j}}\left(P_{i,j} - \hat{Q}_{i,j}\right)\left(\hat{Q}_{i,j} - Q_{i,j}\right).$$

we have

$$-\sum_{j=1}^{m}\frac{P_{i,j}}{\hat{Q}_{i,j}}\left(\hat{Q}_{i,j} - Q_{i,j}\right) = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(\hat{Q}_{i,j} - P_{i,j})^2}{2\hat{Q}_{i,j}} + \frac{(\hat{Q}_{i,j} - Q_{i,j})^2}{2\hat{Q}_{i,j}} - \frac{(Q_{i,j} - P_{i,j})^2}{2\hat{Q}_{i,j}}.$$

Define matrix $B \in \mathbb{R}^{n\times m}$ as $B_{i,j} = M_{i,j}/\hat{Q}_{i,j}$. Using the fact $\hat{Q}_{i,j} \in [\mu_-, \mu_+]$ and result from Lemma 1, we have

$$\frac{1}{2}|P - \hat{Q}|_1 + \frac{|\hat{Q} - Q|_F^2}{2\mu_+} + \varepsilon\langle \partial|Q|_{tr}, \hat{Q} - Q\rangle \leq \frac{|M|_*}{\mu_-}|\hat{Q} - Q|_{tr} + \frac{|P - Q|_F^2}{2\mu_-}.$$

We write the Singular value decomposition of Q as

$$Q = \sum_{i=1}^{r}\sigma_i \mathbf{u}_i \mathbf{v}_i^\top, \tag{1}$$

where r is the rank of Q, σ_i is the i-th singular value of Q, and $(\mathbf{u}_i, \mathbf{v}_i)$ are the left and right singular vectors of Q. Let $U_\perp \in \mathbb{R}^{n\times(n-r)}$ and $V_\perp \in \mathbb{R}^{m\times(m-r)}$ be the orthogonal bases complementary to U and V, respectively. Define the linear operators $\mathcal{P}_Q$ and $\mathcal{P}_Q^\perp$ as

$$\mathcal{P}_Q(Z) = UU^\top Z + ZVV^\top - UU^\top ZVV^\top, \;\; \mathcal{P}_Q^\perp(Z) = Z - \mathcal{P}_Q(Z).$$

According to (1), the subgradient $\partial|Q|_{tr}$ is given by the set $\mathcal{W}$

$$\mathcal{W} = \left\{UV^\top + U_\perp W V_\perp \;:\; W \in \mathbb{R}^{(n-r)\times(m-r)}, \, |W|_* = 1\right\}.$$

Thus by choosing an appropriate matrix W for the subgradient $\partial|Q_{tr}|$, we have

$$\langle \partial|Q|_{tr}, \hat{Q} - Q\rangle \geq -|\mathcal{P}_Q(\hat{Q} - Q)|_{tr} + |\mathcal{P}_Q^\perp(\hat{Q} - Q)|_{tr}$$

and therefore

$$\begin{aligned}&\frac{1}{2}|P-\hat{Q}|_1+\frac{|\hat{Q}-Q|_F^2}{2\mu_+}+\varepsilon|\mathcal{P}_Q^{\perp}(\hat{Q}-Q)|_{tr}\\ \leq\ &\varepsilon|\mathcal{P}_Q(\hat{Q}-Q)|_{tr}+\frac{|M|_*}{\mu_-}|\hat{Q}-Q|_{tr}+\frac{|P-Q|_F^2}{2\mu_-}.\end{aligned}$$

Using the fact

$$\varepsilon\geq\frac{1}{\mu_-}|M|_*,$$

we have

$$|P-\hat{Q}|_1+\frac{|\hat{Q}-Q|_F^2}{\mu_+}\leq 4\varepsilon|\mathcal{P}_Q(\hat{Q}-Q)|_{tr}+\frac{|P-Q|_F^2}{\mu_-}.$$

We consider two cases. In the first case, we assume

$$|P-\hat{Q}|_1\leq\frac{1}{\mu_-}|P-Q|_F^2,$$

in which the bound in theorem trivially holds. In the second case, we have the opposite

$$|P-\hat{Q}|_1>\frac{1}{\mu_-}|P-Q|_F^2,$$

which implies

$$\frac{|\hat{Q}-Q|_F^2}{\mu_+}\leq 4\varepsilon|\mathcal{P}_Q(\hat{Q}-Q)|_{tr},$$

and therefore

$$|\mathcal{P}_Q(\hat{Q}-Q)|_{tr}\leq 4\varepsilon r\mu_+.$$

We complete the proof by plugging the above bound.

1.4 Proof of Theorem 4

Proof. Following the same analysis as that for Theorem 2 in the main paper (see Section 1.3 in this supplementary for its proof), we have

$$\sum_{i=1}^m\frac{(p_i-\hat{q}_i)^2}{\hat{\mathbf{q}}_i}\leq\sum_{i=1}^m\frac{z_i}{\hat{q}_i}(p_i-\hat{\mathbf{q}}_i).$$

Using the fact $\hat{\mathbf{q}}_i\in[\mu_-,\mu_+]$, we have

$$|\mathbf{p}_i-\hat{\mathbf{q}}_i|_2^2\leq\frac{\mu_+}{\mu_-}|\mathbf{z}|_2|\mathbf{p}-\hat{\mathbf{q}}|_2,$$

and therefore

$$|\mathbf{p}_i - \hat{\mathbf{q}}|_2 \leq \frac{\mu_+}{\mu_-}|\mathbf{z}|_2.$$

We finally complete the proof by using the fact

$$\sum_{i=1}^{m} \frac{(p_i - \hat{q}_i)^2}{\hat{q}_i} \geq |\mathbf{p} - \hat{\mathbf{q}}|_1.$$

1.5 Proof of Theorem 5

Proof. We will use the Chernoff bound, i.e. $X_1, \cdots, X_{m_*}$ be independent draws from a Bernoulli distribution with $\mathbb{P}(X = 1) = \mu$. We have

$$\mathbb{P}\left(\frac{1}{m_*}\sum_{i=1}^{m_*} X_i \geq (1+\delta)\mu\right) \leq \exp\left(-\frac{\delta^2 \mu m_*}{3}\right),$$
$$\mathbb{P}\left(\frac{1}{m_*}\sum_{i=1}^{m_*} X_i \leq (1-\delta)\mu\right) \leq \exp\left(-\frac{\delta^2 \mu m_*}{2}\right).$$

Using the Chernoff bound, we have, with a probability $1 - 2\exp(-\delta^2 \mu m_*/2)$

$$|X - \mu|^2 \leq \delta^2 \mu^2.$$

By taking the union bound, we have, with a probability $1 - 2e^{-t}$

$$|\mathbf{z}|_2 \leq \sqrt{\frac{t + \log m}{\mu_- m_*}}|\mathbf{p}|_2.$$

2 Statistics about the Refined Datasets

Table 1. Statistics for the datasets used in the experiments. These datasets are not the original datasets but refined according to our setup. Note NUS-WIDE has two types of tags: the one automatically crawled from Flickr and used for model training, and the one manually annotated.

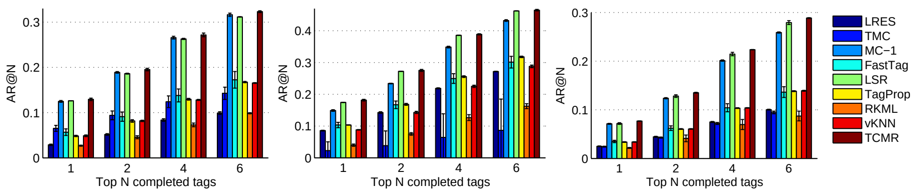
	ESP Game	IAPR TC12	MirFlickr	NUS-WIDE
Number of Images	10,450	12,985	5,231	20,968
Visual feature dimension	1000	1000	1000	500
Vocabulary size	265	291	372	420
Average tags per image	6.41	7.07	5.82	10.4
Min/max tags per image	5/15	5/23	4/43	9/15
Average images per tag	253.0	315.5	81.9	519.6
Min/max images per tag	16/3,439	14/4,752	10/781	78/5,058
Number of observed tags (m_*)*	4	4	3	4

* The number of observed tags when training our proposed model throughout the experimental section if without specific explanation.

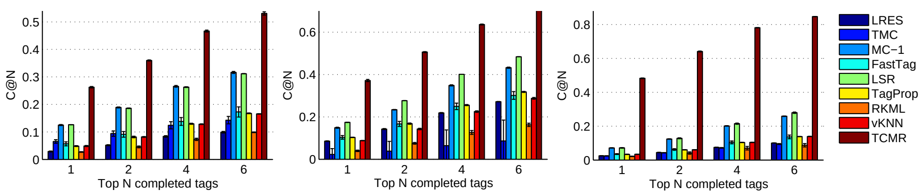
3 Supplementary Experimental Results

In this section, we further present the experimental results of our proposed TCMR in comparison with the baseline approaches.

3.1 Comparison to the state-of-the-art Tag Completion Methods



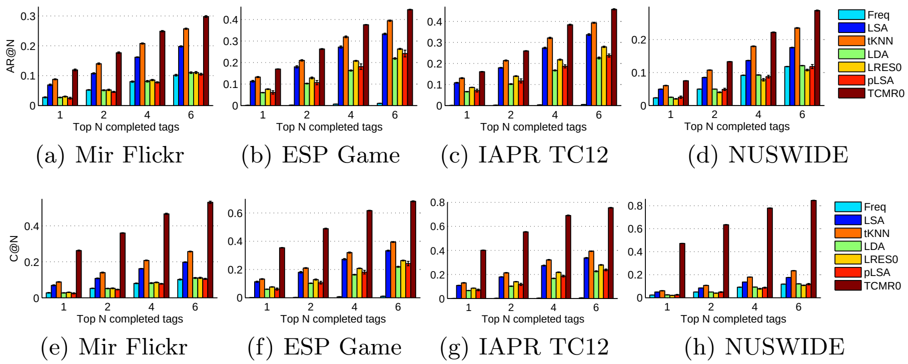
(a) AR@N on Mir Flickr (b) AR@N on ESP Game (c) AR@N on NUSWIDE



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

Fig. 1. Tag completion performance of the proposed method and state-of-the-art baselines on Mir Flickr, ESP Game and NUS-WIDE datasets, reported by $AR@N$ and $C@N$. This figure can be viewed as supplemental to Fig. 1 in the main paper.

3.2 Evaluation of Noisy Matrix Recovery

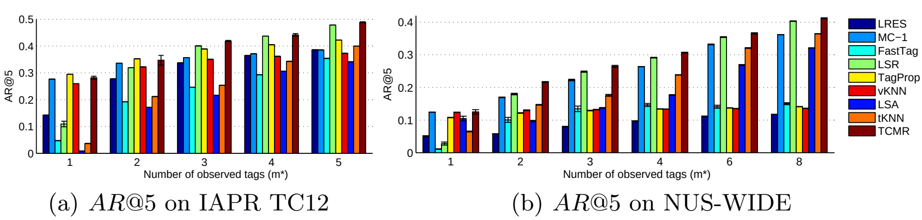


(a) Mir Flickr (b) ESP Game (c) IAPR TC12 (d) NUSWIDE

(e) Mir Flickr (f) ESP Game (g) IAPR TC12 (h) NUSWIDE

Fig. 2. Comparison of different topic models and matrix completion algorithms without taking into account the visual feature. The top row is evaluated by $AR@N$, and the bottom row is evaluated by $C@N$. This figure can be viewed as supplemental to Fig. 2 in the main paper.

3.3 Sensitivity to the Number of Observed Tags



(a) $AR@5$ on IAPR TC12 (b) $AR@5$ on NUS-WIDE

Fig. 3. Tag completion performance with varied number of observed tags, with $AP@5$ reported. This figure can be viewed as supplemental to Fig. 3 in the main paper.