**iTag: Incentive-Based Tagging**

Siyu Lei, Xuan S. Yang, Luyi Mo, Silviu Maniu, Reynold Cheng

Department of Computer Science, The University of Hong Kong

---

### Introduction

- **CollaborativeTagging Data:** facilitate many applications

- **Resource needs sufficient number** of posts to get high-quality tag data.

- **Under-Tag and Over-Tag in Collaborative Tagging Systems**

---

### Approach

**Tag Quality**

- **Relative Frequency Distribution (rfd)** $F_i(k)$:
  Normalized number of occurrence of each tag, after resource $r_i$ has $k$ posts.

- **Stability** $m_i(\omega, k)$:
  Average similarity of rfds’ within window $[k - \omega, k]$.

- **Stable Point:** When stability score surpasses a threshold $T$.

- **Over –Tagging:** Posts given to resources that has passed stable point.

- **Tag Quality:**
  - For resource $r_i$: $q_i(k)$ defined on stability score.
  - For resource set $R$: $q(R, k) = \frac{1}{n} \sum_{i=1}^{n} q_i(k)$

---

**Incentive-Based Tagging**

- **Intuition:** Find the optimal ordering of the resources to achieve the best tagging quality.

- **Input:** A set of tagged resources and budget.

- **Output:** Incentive Allocation.

- **Objective:** Maximize Tag Quality.

- **Optimal Solution:**
  - Dynamic Programming
  - Need to know the posts in the future.

---

### Overview

**Incentive Allocation Strategies**

- **Random (R):** Randomly allocate resources to taggers to tag.

- **Fewest Post First (FP):**
  Prioritize the under-tagged resources.

- **Most Unstable First (MU):**
  Prioritize the most unstable resources; window size $\omega$.

- **Hybrid (FP-MU):**
  FP first, switch to MU when each resource has $\omega$ posts.

---

**Results**

- **Dataset:** 5000 urls and their posts from del.icio.us

- **FP & FP-MU** close to optimal;
  - $FC$ hardly increases the quality.

- **50%** of the posts by $FC$ are over-tagging.

---

**Conclusion:** iTag can significantly improve tag data quality for providers with the least amount of money.

---

### Contact

Siyu Lei  
sylei@cs.hku.hk

---

**Reference:** X. Yang, R. Cheng, L. Mo, B. Kao, and D. Cheung