Identifying Malicious Players in GWAP-based Disaster Monitoring Crowdsourcing System

Changkun Ou, Yifei Zhan
Institute of Computer Science
University of Munich
changkun.ou@lmu.de
yifei.zhan@campus.lmu.de

Yaxi Chen
The Key Laboratory for Computer Systems of State Ethnic Affairs Commission
Southwest Minzu University
yaxichen@swun.cn

ICAIBD' 19, Chengdu, China
May 26, 2019

Outline

1. Background & Motivation
2. Preliminaries
3. Main Results
   - Player Rating Model (PRM)
   - Disaster Evaluation Model (DEM)
   - Model Initialization
4. Evaluation & Discussion
   - Simulation
   - Limitations
5. Conclusions

Outline

Background: Human Computation in 1 Minute

What are **Human Computation** systems?

- "Systems that combine humans and computers to solve large-scale problems that neither can solve alone" (Luis von Ahn, retrived 30 Apr. 2019)
- Software systems with humans in the loop, human as explicit (or active) or implicit (or passive) contributors

Background: Human Computation in 1 Minute

What are **Human Computation** systems?

- "Systems that combine humans and computers to solve large-scale problems that neither can solve alone" (Luis von Ahn, retrived 30 Apr. 2019)
- Software systems with humans in the loop, human as explicit (or active) or implicit (or passive) contributors

Human Computation systems can be seen as Crowdsourcing markets (Wisdom of crowds). Useful inputs (wisdom) can be gained from a group of persons provided: **Diversity of opinion; Idependence; Decentralization; Aggregation.** (James Surowiecki, 2005)

Background: Human Computation in 1 Minute

What are **Human Computation** systems?

- "Systems that combine humans and computers to solve large-scale problems that neither can solve alone" (Luis von Ahn, retrived 30 Apr. 2019)
- Software systems with humans in the loop, human as explicit (or active) or implicit (or passive) contributors

Human Computation systems can be seen as Crowdsourcing markets (Wisdom of crowds). Useful inputs (wisdom) can be gained from a group of persons provided: **Diversity of opinion; Idependence; Decentralization; Aggregation.** (James Surowiecki, 2005)

Game-With-A-Purpose (GWAP) tries to hide actual intent away from players and aggregates human inputs for solving difficult problems.

Motivation

- Non-profit organizations (e.g. UNICEF) has lack of resources in monitoring disaster regions, an automated system is essential.

Motivation

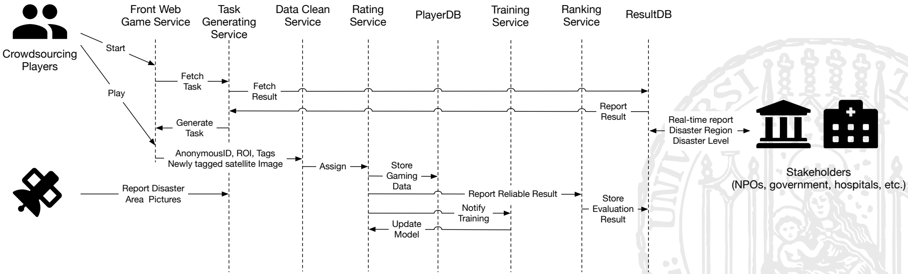
- Non-profit organizations (e.g. UNICEF) has lack of resources in monitoring disaster regions, an automated system is essential.
- Sucessful disaster monitoring requires
  - *reliable predictions*: system and algorithm design
  - *low costs maintains*: GWAPs-based crowdsourcing
- *Malicious player detection* is critical in disaster monitoring and guarentees the health of a GWAP-based human computation system.

Outline

System Architecture

The system consist of three components:

- **task generating service**
- **rating service**
- **ranking service**



System Interface

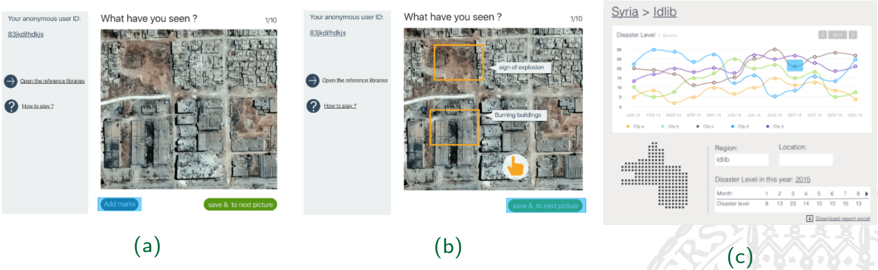


Figure: System interface. a) Player game panel overview; b) Multi-tags selection for selected areas; c) Disaster level report in stakeholder view.

Preliminaries

Definition (Region of Interests, ROI)

An ROI represents a subset of $\mathbb{R}^2$. The i-th ROI from player p in image k is denoted by $ROI_{p,i,k}$.

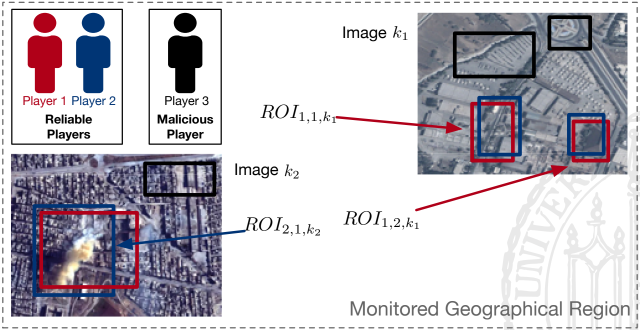


Figure: Reliable players (red and blue) draw rectangles to indicate area with disaster, however malicious player does not cooperate (black) selects other ROIs.

Preliminaries (cond.)

Definition (Tag Vector, TV)

Assuming n different tags g_1, g_2, ..., g_n for a certain image k, the tag vector is defined by $\mathbf{T}_{p,i,k} = (|g_1|, |g_2|, ..., |g_n|)^\top$ of $ROI_{p,i,k}$ where g_l is the l-th tag where $l = 1, 2, ..., n$, $|g_l|$ is the count of g_l in a player task object, and n equals to the number of tags.

Outline

Player Rating Graph (PRG)

Definition (System Weight Vector)

For n different tags g_1, g_2, ..., g_n. Let $|g_i|$ is the count of g_i in the system. A system weight vector $\mathbf{v} = (p(g_1), p(g_2), ..., p(g_n))^\top$, where

$$p(g_i) = \frac{|g_i|}{\sum_{j=1}^{n} |g_j|}, i = 1, ..., n. \tag{1}$$

Lemma (Properties)

$p(g_i)$ holds the properties:

- $0 \leq p(g_i) \leq 1$
- $\sum_{i=1}^{n} p(g_i) = 1$
- $\sum_{i=1}^{s} p(g_{r_i}) \leq 1$

Player Rating Graph (PRG) cond. I

Definition (Image Weight Vector)

For different tags g_{r_1}, g_{r_2}, ..., g_{r_s} in a certain k, the image weight vector is a vector for image k that is composed by part of the system weight vector where $\mathbf{v}_k = (p(g_{r_1}), p(g_{r_2}), ..., p(g_{r_s}))^\top$ with $r_i(i = 1, 2, ..., s) \in \{1, 2, ..., n\}$, $r_i \neq r_j(i \neq j, j = 1, 2, ..., s)$ and $s \leq n$.

Player Rating Graph (PRG) cond. II

Definition ((Asymmetric) Players ROI Matching Ratio, PRMR)

For player p q, and a certain image k:

$$\mathrm{PRMR}(p, q, i, j, k) = \frac{|ROI_{p,i,k} \cap ROI_{q,j,k}|}{|ROI_{p,i,k}|} \tag{2}$$

where $ROI_{p,i,k}$ is the i-th selected ROI from player p, and $|ROI_{p,i,k}|$ is the surface area of $ROI_{p,i,k}$.

Lemma (PRMR Bounds)

The inequality holds:

$$0 \leq PRMR(p, q, i, j, k) \leq 1 \tag{3}$$

Player Rating Graph (PRG) cond. III

Definition ((Asymmetric) Player Input Tag Correlation, PITC)

For two different tag vectors $\mathbf{T}_{p,i,k}, \mathbf{T}_{q,j,k}$ from player p, q image weight vector $\mathbf{v}_k$, PITC is defined as follows:

$$\text{PITC}(p, q, i, j, k) = \frac{Cov(\mathbf{T}_{p,i,k}, \mathbf{T}_{q,j,k}; \mathbf{v}_k)}{Cov(\mathbf{T}_{p,i,k}, \mathbf{T}_{p,i,k}; \mathbf{v}_k)} \tag{4}$$

where $Cov(\mathbf{X}, \mathbf{Y}; \mathbf{w})$ is the weighted covariance of $\mathbf{X}$ and $\mathbf{Y}$.

Lemma (PITC Bounds)

The inequality holds:

$$-1 \leq PITC(p, q, i, j, k) \leq 1. \tag{5}$$

Player Rating Graph (PRG) cond. IV

Definition (PRG Edge Weight)

For a image k, the weight of the PRG between player p and q is:

$$w_{p,q,k} = \sum_{j=1}^{n} \sum_{i=1}^{m} \mathrm{PRMR}(p,q,i,j,k)\,(\mathrm{PITC}(p,q,i,j,k) + 2) \tag{6}$$

where p selected m ROIs and q selected n ROIs.

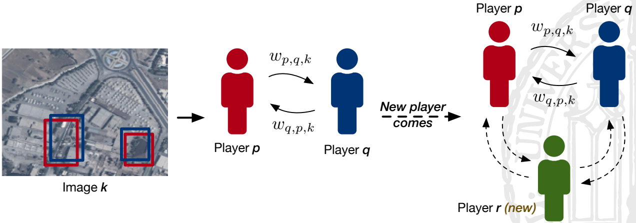


Figure: PRG for certain images: Assume player p and q are former reliable players. A new player is composed with former players in the graph as a game network.

Player Rating Graph (PRG) cond. V

Let a normalized adjacency matrix calculated as follows:

$$\mathbf{A}_k = (a_{p,q,k}) = \left(\frac{w_{p,q,k}}{\sum_q w_{p,q,k}}\right) \tag{7}$$

where k is the image indicator. We have

Theorem (Soundness)

The normalized adjacency matrix A_k of PRG of a certain image k is irreducible, real, non-negative, and column-stochastic, with positive diagonal element.

Player Rating Graph (PRG) cond. VI

According to Perron-Frobenius theorem, one can infer that there exists an uniqueness eigenvector $\mathbf{V}_k = (\lambda_{1,k}, ..., \lambda_{n,k})^\top$ of $\mathbf{A}_k$ (Perron vector), with an uniqueness eigenvalue $\rho(\mathbf{A}_k)$ is the spectral radius of $\mathbf{A}_k$ (Perron root), such that:

$$\mathbf{A}_k \cdot \mathbf{V}_k = \rho(\mathbf{A}_k) \cdot \mathbf{V}_k, \lambda_{i,k} > 0, \sum_{i=1}^{n} \lambda_{i,k} = 1.$$

Definition (Trust Value, λ)

A trust value $\lambda_{i,k}$ of player i on image k is a score that equals to the i-th component of the Perron vector of the normalized PRG adjacency matrix $\mathbf{A}_k$.

Malicious Detection Algorithm

Algorithm 1: Malicious Player Detection

input : New Player p, Reliable Player $p1, p2, \ldots, p_m$,
Task Images $k_1, k_2, \ldots, k_{2n}$, Acceptance Threshold δ
output: Reliability of Player p
begin
    counter $\longleftarrow$ 0
    reliability $\longleftarrow$ false
    for $k \in [k_1, k_2, \ldots, k_{2n}]$ **do**
        if *k is tagged image* **then**
            calculate $\lambda_{p,k}, \lambda_{p_1,k}, \ldots, \lambda_{p_m,k}$
            if $\lambda_{p,k} \geq \frac{1}{m} \sum_{i=1}^{m} \lambda_{p_i,k}$ **then**
                counter $\longleftarrow$ counter $+1$
            end
        end
    end
    if *counter* $\geq \delta$ **then**
        reliability $\longleftarrow$ true
    end
end

The *acceptance threshold* is a hyperparameter that can be set beforehand. For instance, if $\delta = 1$, the new player only needs to pass one singular image of all tagged images; if $\delta = n$ (half images of the task), the new player has to pass all tagged images, which makes the system unbreakable if the system is initialized by a trusted group.

Malicious Detection Algorithm (Cond.)

New player carries new tags into the system will influence the tag vector calculation and cause the weight not computable due to the inequal dimensions of the tag vector of new player and old player. Solution:

- If a new player does not provide new tag: Directly perform the calculation with the algorithm;
- If a new player carries new tags only: Directly drop them because they are unreliable;
- If a player carries both selected and new tags: a) Perform the calculation with the algorithm without new tags; b) Merge and update all weight vector v via formula 6 if the player is reliable; c) Otherwise drop and mark the result as unreliable.

Disaster Evaluation Model (DEM)

Definition (Disaster Level Δ)

A monitor region is composed by images $k_1, ..., k_n$. Each image exists r_{k_i} number of ROIs with $i = 1, ..., n$, and each ROI is tagged with tags $g_1, ..., g_m$. The *disaster level* Δ of a monitor region is:

$$\Delta = \sum_{j=1}^{m} \left(p(g_j) \frac{\sum_{g_j} |ROI|}{\sum_{i=1}^{n} |k_i|} \right) \tag{8}$$

where $|ROI|$ is the surface area of a ROI, $\sum_{g_j} |ROI|$ means accumulated surface area of all ROIs that tagged by g_j, and $|k_i|$ is the surface area of image k_i.

Theorem (Denseness)

The disaster level Δ is dense in internal $[0, 1]$.

Determine The Size of Trusted Group

The PRM is based on graph centrality calculation, which means we need a (at least) two dimensional matrix to perform the overall model calculation. Hence, with the new player, *the minimum number of the initial trusted group is 1*. Then the initial trusted group (one person) with the new player form a two dimensional adjacency matrix that makes the model computable. For larger initial trusted groups, the trust value can be simply initialized to $\frac{1}{n}$ with n is the number of initial trusted group.

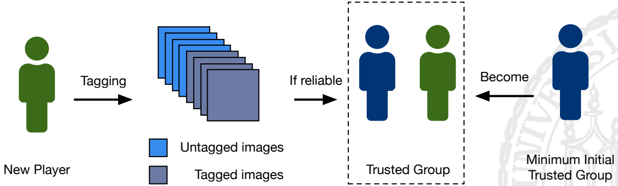


Figure: Initialization of PRM

Outline

Simulated Evaluation

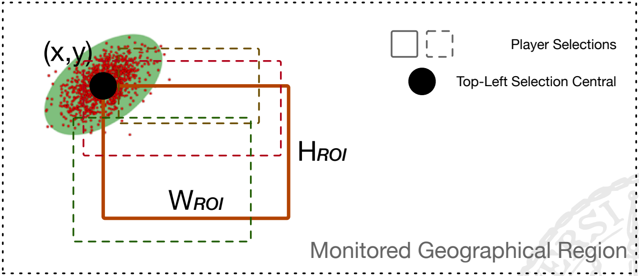


Figure: An example of ROI simulation which can be used in the system evaluation.

Data Leakage and Information Loss

Figure: Information loss may occur on the intersection lines; a possible solution is to perform a "half shifting" cut.

Limitations

Outdated Evaluation If none of the new images gets evaluated, then the disaster level will not be updated.
Solution: time series prediction.

Game Playability Players may meet the situation that there is no available ROI in several continuous rounds.
Solution: pre-filtering.

Outline

Take Away

- Human-computation systems solve problems that neither computer or human can solve alone.
- We proposed Player Rating Graph Model and Disaster Evaluation Model and mathematically proved its soundness and completeness.
- Our models solves the model initialization problems in human computation field.
- The models are generic and can be easily apply to any other similar systems.
- Simulation and Half Shifting cut are proposed for evaluation and data security.
- Time series prediction and image pre-filtering are proposed to address outdated evaluation and game playability for our future works.

Thank you for your attention!