Truth discovery in crowd sensing

Problem

You have a set of volunteers(sensors) to give their information for a set of entities. Example: image labling.

- **Input:** Conflicting information about the same set of entities provided by multiple sensors.
  - Each information of a provider i to an entity j is labeled as $\vec{x}_i^j$
  - Not all provider observes all entity, so we have an observation matrix A
- **Goal:** Discover trustworthy information (i.e., the truths) from conflicting data on the same entity $\vec{x}_i^*$
- **Challenge:** The **reliability** of each sensor should be taken into account when aggregating the observations. However, the reliability information is usually unknown a priori

> Hint: use a variable w to represent the *reliability* of a provider.

Optimization

$$\mathbf{P}: \min_{\{X,W\}} \sum_{i=1}^{t} \sum_{j=1} a_{ij} w_j \left\| \vec{x}_i^* - \vec{x}_i^j \right\|^2$$

$$\text{s.t.} \quad \sum_{j=1}^{n} \exp(-w_j) = 1$$

$$\vec{x}_i \geq \vec{0}, \quad |\vec{x}_i| = 1 \text{ for } i = 1, 2, \ldots, t$$

Algorithm

This problem is non-convex

Fix w -> $\vec{x}$ is convex problem(LS?)

Fix $\vec{x}$ -> w is convex problem (Affine)

Thus, the algorithm is like EM and coordinate descends.

$$\vec{x}_i^{(t+1)} = \frac{\sum_j a_{ij} w_j \vec{d}_i^j}{\sum_j a_{ij} w_j}$$

$$w_j^{(t+1)} = \log \frac{\sum_j \sum_i a_{ij} \left\| \vec{x}_{i.}^{(t+1)} - \vec{d}_{i.}^j \right\|^2}{\sum_i a_{ij} \left\| \vec{x}_i^{(t+1)} - \vec{d}_i^j \right\|^2}$$

The truth discovery algorithm converges within finite number of iterations. But it may converges to the **local minimum**.

Discussion

The problem of $\mathbf{min}_{w,x} \sum_k w_k \|x_k - x\|_2^2$ can be regarded as minimizing some entropy \

> $\mathbf{min} -p \log p$ where the $p_k = \frac{\|x_k - x\|_2^2}{\sum_k \|x_k - x\|_2^2}$

Reference

Lu Su, Q. Li, S. Hu, S. Wang, J. Gao, H. Liu, T. Abdelzaher, J. Han, X. Liu, Y. Gao, L. Kaplan, “Generalized Decision Aggregation in Distributed Sensing Systems”, in RTSS 2014.