Online Influence Maximization under Linear Threshold Model

Shuai Li, Fang Kong, Kejie Tang, Qizhi Li, We Chen

Motivation

- Online influence maximization (OM): A sequential decision-making problem
- Selects a set of users and provides them with free products
- Receives feedback from the information diffusion process

Common influence propagation models: independent cascade (IC) model, linear threshold (LT) model

Existing works mainly focus on the IC model with edge-level feedback

The LT model assumes the information spreads through each edge independently.

The edge-level feedback mechanism the learner could observe the influence status of each edge once its start node is influenced.

The LT model characterizes the herd behavior that often occurs in real information diffusion processes, that with more active in-neighbors, a user becomes much more likely to be influenced.

Setting

Graph $G = (V,E)$: $V$ is the set of users (nodes) and $E$ is the set of relationships (edges)

Each edge in $G$ is associated with a weight $w_{uv}$ representing the influence of user $u$ on $v$.

$n = |V|, m = |E|$: Node number, edge number, and the diameter respectively, where the diameter of the graph is defined as the maximum (directed) distance between the pair of nodes in any connected component.

The diffusion process starts from seed set $S$ under LT model

Each node is assigned with a threshold $\theta_v$, which is independently drawn from $[0, 1]$ and characterizes the susceptibility level of $v$.

Let $S_t$ be the set of activated nodes by the end of time $t$.

At time $t = 0$, only nodes in the seed set are activated: $S_0 = S$.

At time $t + 1$ for any node $v \notin S_t$ that has not been activated yet, it will be activated if the aggregated influence of its active in-neighbors exceeds its threshold: $\sum_{u \in N(v) \cap S_t} w_{uv} \geq \theta_v$.

Such diffusion process will last at most $D$ time steps.

The size of the influenced nodes: $|S_t(S, \theta)| = |S|$

Let $\rho_t(S, w) = \mathbb{E}[|S_t(S, w)|]$ be the influence spread of seed set $S$ where the expectation is taken over all random variables $\theta_v, \forall v \in V$.

The maximum influence spread under weight vector $w$ with limited size $K$

Some notations in the influence process

- $\tau_{\text{init}}(v) = \min \{0, \ldots, D : N(v) \cap S_t \neq \emptyset\}$ is the earliest time step when node $v$ has active neighbors, set $\tau_{\text{init}}(v) = D + 1$ if node $v$ has no active in-neighbor until the diffusion ends.

- $\tau(v) = \tau_{\text{init}}(v) + \tau_{\text{delay}}(v)$, where $\tau_{\text{delay}}(v) = \min \{u / v \in N(v) \cap S_t \}$ is the time step when node $v$ is influenced, set $\tau(v) = D + 1$ if node $v$ is finally not influenced after the information diffusion ends.

- $E_{\tau(v)} = \{u / \tau(v) \leq \tau(u) \leq D\}$ is the set of incoming edges associated with active in-neighbors of $v$ at time step $\tau(v)$.

The online IM (OM) problem, in each round $t$:
- The learner chooses a set of nodes $S_t$ with limited size $K$
- The learner observes (full) node-level feedback $S_t, D_1, D_2, \ldots, D_t$, where $S_t$ represents the set of active nodes by time step $t$ in $|D_t|$ in the diffusion process in this round.

The learner updates its knowledge on unknown weights using the observed feedback, which helps the seed set selection in the next round.

The goal of the learner is to minimize the expected cumulative $\gamma$-scaled regret

$R(T) = \mathbb{E} \left[ \sum_{t=1}^{T} \gamma_t \cdot \tau(v) \mid \tau(v) \notin S_t \right] = \mathbb{E} \left[ \sum_{t=1}^{T} \tau(v) \cdot \gamma_t \mid \tau(v) \notin S_t \right]
$

Figure 1: An example of diffusion process starting from $S = \{v\}$ under LT. The upper part describes an influence diffusion process where yellow nodes represent influenced nodes by the current time. The lower part describes what $E_{\tau(v)}, E_{\tau(v) - 1}$ are where we use blue (red) color to represent the edges and the associated active in-neighbors in $E_{\tau(v)}, E_{\tau(v) - 1}$ respectively for the objective black node.

Algorithm

LT-ListIC Algorithm:

1. Input: Graph $G = (V,E)$, seed set cardinality $K$, exploration parameter $\rho > 0$ for any $t, v$, offline oracle PairGracels
2. Initialize: $M_0, \gamma_0 \leftarrow 0 \in \mathbb{R}^{|V|\times|V|}, \lambda_0 \leftarrow 0 \in \mathbb{R}^{N(t)}, \gamma_0 \leftarrow 0 \in \mathbb{R}^{|V|}$ for any node $v \in V$
3. for $t = 1, 2, 3, \ldots,$
4. Compute the confidence ellipsoid $C_t = \left\{ \gamma' \in [0,1]^{N(t)} : \|\gamma' - \gamma(t)\|_{1,2} \leq \rho \gamma(t) \right\}$
5. for any node $v \in V$
6. Compute the pair $(S_t, \psi)$ by PairGracels with confidence set $C_t = \{C_t\}_{t=1}^{T}$ and seed set cardinality $K$
7. Select the seed set $S_t$ and observe the feedback
8. $\delta$Update
9. for node $v \in V$
10. Initialize $A_t(v) = 0 \in \mathbb{R}^{N(t)}, \lambda_t(v) = 0 \in \mathbb{R}$
11. Uniformly randomly choose $r \in \{0, 1\}, \tau_r(v) \equiv \min(t \mid \tau_r(v) \leq \tau_r(v) - 1)$
12. if $v$ is influenced and $r = \tau_r(v) - 1$
13. else if $\tau_r(v) \neq \tau_r(v)$ or $\tau_r(v) - 1$ but $v$ is not influenced
14. $\lambda_t(v) \leftarrow \gamma_r(v), \psi_t(v) = 0$
15. $M_t(v) \leftarrow M_t(v) = A_t(v) + \gamma_t(v) \psi_t(v)$

Analysis

- For the seed set $S$, define the set of all nodes related to a node $v$, $V_{S,v}$, to be the set of nodes that are on any path from $S$ to $v$ in graph $G$.
- For seed set $S$ and node $v \in V \setminus S$, define $N_{S,v} = \{w / w \in V_{S,v}\} \leq n - K$ to be the number of nodes that $v$ is related to.
- Then for the vector $\gamma_t(S,v) = |N_{S,v}|$, define the upper bound of its $L_2$-norm over all feasible seed sets

$\gamma(G) = \max_{S \subseteq V} \sum_{v \in S} |N_{S,v}| \leq (n - K)\sqrt{\pi} = O(k^{1/2}),$

which is a constant related to the graph.

Theorem 1 (GOM bounded smoothness): For any two weight vectors $w, w' \in [0,1]^n$ with $\sum_{v \in V} w(v) - \gamma_t(S,v) \leq 1$, the difference of their influence spread for any seed set $S$ can be bounded as

$|\gamma(S,v') - \gamma(S,v)| \leq \mathbb{E} \left[ \sum_{v \in S} \sum_{w \in \mathcal{W}} (w(v') - w(v)) \left| \tau(v) - \tau(v') \right| \right]
$

where the definitions of $\tau(v), \gamma_t(v)$ and $E_{\tau(v)}$ are all under weight vector $w$, and the expectation is taken over the randomness of the thresholds on nodes.

Theorem 2 (Upper Bound): Suppose the LT-ListIC runs with an $(\alpha, \beta)$-approximation PairGracels and parameter $\rho < \gamma_t(v) = \sqrt{\log(2) + \log 2 + \sqrt{\pi}}$ for any node $v \in V$.

Then the $\alpha$-scaled regret satisfies

$R(T) \leq 2\alpha\gamma(G)D\sqrt{\ln(T)\log(1/\alpha) + \log(1/\gamma) + \delta} T(n - k).$

When $\delta = \frac{1}{2(v \in V, 1)}$, $R(T) \leq \gamma(G)D\sqrt{\ln(T/\delta)}$ for some universal constant C.

Conclusions

- Formulate the problem of OIM under LT model with node-level feedback and design how to distill effective information from observations.
- Prove a novel OIM bounded smoothness property for the spread function.
- Propose LT-ListIC algorithm with rigorous theoretical analysis and show a competitive regret bound of $O((\log(n)\sqrt{T} + \log T))$.
- Design GOM-ETC algorithm with theoretical analysis on its distribution-dependent and distribution-independent regret bounds.
- Efficient, applies to both LT and IC models, and has less requirements on feedback and offline computation.
- Future work: The OIM problem under IC model with node-level feedback, Applying Thompson sampling to influence maximization.

References