Lecture 10: Network Influence Seed Optimization

Xiaotie Deng

AIMS Lab
Department of Computer Science
Shanghai Jiaotong University

November 22, 2017

Social Network Influence

Posting Forward in Social Networks

- Friend Relationships
- Influential Nodes in Social Networks (Page Rank)
- Restaurants, movies, tourist locations and commercial products.
- Different people may have different influences.

Viral Advertisement

- One person's activity may have an impact on her/his friends.
- Seed the right ad on the right person may create a cascading over friends and friends.
- How to choose a minimum subset of people in social network to have the maximum effect of influence.
- A model of information epidemics

Independent Cascade Model

- Directed graph $G = (V, E; p, A_0)$.
  - $A_0 \subseteq V$ is the initially activated seed set.
  - $p : E \rightarrow Q_+$ where $p(i,j)$ is the independent state transition probability $P(j = 1 | i = 1)$ of influence of i on j.
  - If i is activated, j will be influence to become active with probability $p(i,j)$.
- Each activated node initiate an activates of its neighbors only once. Whether it succeeds or not, it makes no further effort.
- Algorithmic Issue: How to find the best seed set of a given size?

Linear Threshold Model

- Given an initial active set A_0,
  - a given set of weight $b_{v,w}$ such that $\sum_{w\in\Gamma(v)} b_{v,w} \leq 1$ is given.
  - a randomly chosen vertex specific threshhold θ_v is generated
- Deterministic diffusion steps then follow:
  - Step t: all nodes inherit their past activeness.
  - Each node v with $\sum_{w\in\Gamma(v),w \text{ active at time } t-1} b(v,w) \geq \theta_v$ become active.

Optimal Choice of Initial Active Set

- Given an initial active set A_0,
  - NP-hard to find the exact optimal solution.
  - $(1 - \frac{1}{e} - \epsilon)$-approximation by a greedy hill-climbing strategy.

Submodular function optimization

Monotone Submodular Function

- $f(S \cup \{v\}) \geq f(S)$
- $\forall S \subseteq T : f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$
- Finding the optimal seed set is NP-hard.
- Greedy hill climbing achieves $(1 - \frac{1}{e})$-approximation.
  - Repeatedly add an element which gives the maximum marginal gain.

Prove Submodular Property

- Prove the resulting influence function $\sigma(\cdot)$ is submodular.
- Difficulty to apply it here: exact greedy element cannot be found.
- A $(1-\epsilon)$ of the optimum is designed to achieve an $1-\frac{1}{e}-\epsilon$ factor approximation of the influence problem.

Independent Cascade

- Consider $\sigma(A \cup \{v\}) - \sigma(A)$ by an equivalent view.
  - $\forall e = (w, v) \in E$ a coin of bias $p_{w,v}$ is flipped in the beginning (and read whenever needed).
  - Claim: A node x ends up active iff there is a path from some node in A_0 to x consisting of live edges.
- Influence function is submodular.
  - Given coin flipping X, $\sigma_X(A) = \cup_{v \in A_0} R(v, X)$. where $R(v, X)$ is the set of vertices reachable from A_0 given flipping X.
  - $\sigma_X(S \cup \{v\}) - \sigma_X(S)$ is the number of elements in $R(v, X)$ that are not already in $\cup_{u \in S} R(u, X)$
  - It follows that $S \subseteq T$ implies $\sigma_X(S \cup \{v\}) - \sigma_X(S) \geq \sigma_X(T \cup \{v\}) - \sigma_X(T)$
- As $\sigma(A)$ is a convex combinations $\sigma_X(A)$, the claim follows.

Independent Cascade is NP-hard

- Reduction from set cover.