Submodular and Influence Maximization

Mauro Sozio

Telecom ParisTech

January 21, 2019

Submodular Functions

Definition 1 (Submodular Function)

Given a finite set V, a function $f : 2^V \to \mathbb{R}$ is submodular if it satisfies any of the following equivalent definition:

- $f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$, for any $v \in V$, and for every S, T such that $S \subseteq T$.
- for every $S, T \subseteq V$, $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$.

Example of a submodular function?

Submodular Maximization

Definition 2 (Submodular Maximization)

Given a set V, find a set $S \subseteq V$ such that $|S| \leq k$ and $f(S)$ is maximum.

It is NP-Hard (contains Hitting Set as a special case).

It can be approximated with a greedy algorithm in some cases of interest.

Greedy Hill-climbing Algorithm

Start with the empty set and repeatedly add an element v with maximum marginal gain $f(S \cup \{v\}) - f(S)$, where S is the current set.

Theorem 3 (Approximation Guarantee of HC [2])

If f is monotone non-negative and submodular, the Hill-climbing algorithm provides a $(1 - 1/e)$-approximation.

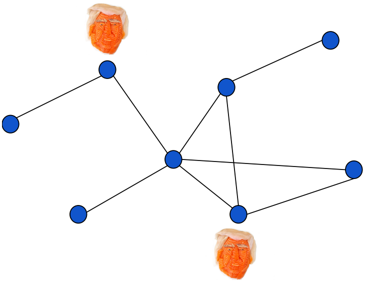
Spread of Ideas/Innovations in Social Networks

- An idea or innovation appears and it might either die out quickly or spread through the population.
- Examples: the use of smartphones among college students, the adoption of a new drug within the medical profession, or the spread of the message "Drumpf is a liar and a dangerous president"[1].

[1] Any similarity with fictitious events or characters was purely coincidental

Spread of Ideas/Innovations in Social Networks

- An idea or innovation appears and it might either die out quickly or spread through the population.
- Examples: the use of smartphones among college students, the adoption of a new drug within the medical profession, or the spread of the message “Drumpf is a liar and a dangerous president”[1].



[1] Any similarity with fictitious events or characters was purely coincidental

Influence Maximization

Natural questions are:

1. will the message/innovation spread or die out quickly? See also [5].
2. can we target a set of users in the network so as to maximize the spread? E.g. we wish to advertise a new smartphone model to a set of "influential" users in the network.

We will deal with the second problem: we wish to find a set of at most k users so as to maximize the spread in the social network.

We need a diffusion model...

Diffusion Models: Linear Threshold

Given an undirected graph $G = (V, E)$ [2], weights $b_{v,w} \geq 0$ for every edge vw such that $\sum_{w \in \Delta(v)} b_{v,w} \leq 1,\ \forall v \in V$, a set $A_0 \subseteq V$ of active nodes the diffusion process unfolds as follows:

1. for every node v choose θ_v uniformly at random from $[0, 1]$.
2. At each step t, $t = 1, \ldots, n$ let S_t be the set of nodes v such that

$$\sum_{w \in A_{t-1} \cap \Delta_v} b_{v,w} \geq \theta_v .$$

Then, $A_t := A_{t-1} \cup S$. Stop as soon as $A_t = A_{t+1}$.

[2] generalization to directed graphs is straightforward

Diffusion Models: Linear Threshold

Intuition:

- the threshold θ_v represents the latent tendency of node v to adopt the innovation (e.g. smartphone) when their neighbors (friends) do.
- θ_v's are randomly chosen modeling the lack of knowledge of their values. We are in effect averaging over possible values for all nodes.

We shall consider the *Independent Cascade Model* as well ...

Diffusion Models: Independent Cascade

Given an undirected graph $G = (V, E)$, a probability $p_{v,w}$ for every edge vw and a set A_0 of active nodes, the set of active nodes A_t at step t is determined as follows:

- when node v first becomes active, is given a single chance to activate each currently inactive neighbor w with probability $p_{v,w}$ (independently from the past and other edges).
- If v succeeds then w will become active, otherwise v cannot make further attempts to activate w.
- The process runs until $A_t = A_{t+1}$.

Definition 4

Let $\sigma(A)$ be the expected number of nodes that are active at the end of a diffusion process, given that the initial set of active nodes is A.

Theorem 5

For an arbitrary instance of the Independent Cascade Model, the influence function $\sigma(\cdot)$ is submodular.

Theorem 6

For an arbitrary instance of the Linear Threshold Model, the influence function $\sigma(\cdot)$ is submodular.

Independent Cascade Model: Submodularity

Proof of Theorem 5 (Independence Cascade is Submodular).

Let H be the *undirected* graph obtained by sampling each edge vw in $G = (V, E)$ independently, with probability p_{vw}. Given a set A_0 of active nodes, it holds that a node is active at the end of IC iff it is connected with a node in A_0.

Independent Cascade Model: Submodularity

Proof of Theorem 5 (Independence Cascade is Submodular).

Let H be the *undirected* graph obtained by sampling each edge vw in $G = (V, E)$ independently, with probability p_{vw}. Given a set A_0 of active nodes, it holds that a node is active at the end of IC iff it is connected with a node in A_0.

Let $\sigma_H(A)$ be the number of active nodes when $A = A_0$ are initially active and the graph H is sampled. $\sigma_H(\cdot)$ is submodular, in that, $\sigma_H(S)$ is equal to the number of nodes being in the same connected component of S (set cardinality is submodular). Moreover,

Independent Cascade Model: Submodularity

Proof of Theorem 5 (Independence Cascade is Submodular).

Let H be the *undirected* graph obtained by sampling each edge vw in $G = (V, E)$ independently, with probability p_{vw}. Given a set A_0 of active nodes, it holds that a node is active at the end of IC iff it is connected with a node in A_0.

Let $\sigma_H(A)$ be the number of active nodes when $A = A_0$ are initially active and the graph H is sampled. $\sigma_H(\cdot)$ is submodular, in that, $\sigma_H(S)$ is equal to the number of nodes being in the same connected component of S (set cardinality is submodular). Moreover,

$$\sigma(A) = \sum_{H \in \mathcal{G}_n} Pr(H) \cdot \sigma_H(A),$$

where $\mathcal{G}_n$ is the set of all graphs on $|V| = n$ nodes. Then, $\sigma(\cdot)$ is submodular, in that, it is a non-negative linear combination of submodular functions.

Linear Threshold Model: Recall

Given an undirected graph $G = (V, E)$, weights $b_{v,w} \geq 0$ for every edge vw such that $\sum_{w \in \Delta(v)} b_{v,w} \leq 1, \ \forall v \in V$, a set $A_0 \subseteq V$ of active nodes the diffusion process unfolds as follows:

1. for every node v choose θ_v uniformly at random from $[0, 1]$.
2. At each step t, $t = 1, \ldots, n$ let S_t be the set of nodes v such that

$$\sum_{w \in A_{t-1} \cap \Delta_v} b_{v,w} \geq \theta_v.$$

Then, $A_t := A_{t-1} \cup S$. Stop as soon as $A_t = A_{t+1}$.

Linear Threshold Model: Same argument as IC?

Exercise: Let $X = x_1, \ldots, x_n$ be a set of rational numbers in $[0, 1]$. Let $\sigma_X(A)$ be the number of active nodes at the end of the LT process (with initial set A) where $\theta_{v_i} = x_i$, $i = 1, \ldots, n$. Show that there exists an instance of LT and X such that $\sigma_X(\cdot)$ is not submodular.

Therefore we cannot use the same argument we used in the previous proof.

Linear Threshold Model: Submodularity

Proof of Theorem 6 (Linear Threshold is Submodular).

Let A_t be the set of nodes being active at step t. To compute the probability that a node v becomes active at step $t+1$ we can assume that the threshold for v is chosen (randomly) at step $t+1$. This does not change the outcome of the process (*Principle of Deferred Decisions* [1]).

The prob. that v becomes active at the beginning of step $t+1$ is then:

$$Pr(v \in A_{t+1} | v \notin A_t) = \frac{\sum_{u \in A_t \setminus A_{t-1}} b_{v,u}}{1 - \sum_{u \in A_{t-1}} b_{v,u}},$$

in that, it must be $\theta_v \in (\sum_{u \in A_{t-1}} b_{v,u}, \sum_{u \in A_{t-1}} b_{v,u} + \sum_{u \in A_t \setminus A_{t-1}} b_{v,u}]$.

Proof of Theorem 6 (Continue).

Let H be the *directed* graph obtained as follows: Each node v in G chooses exactly one node w with probability $b_{v,w}$ from its neighbors, while it selects no neighbors with probability $1 - \sum_{w \in \Delta_v} b_{v,w}$. Such a neighbor becomes its *incoming* edge.

Proof of Theorem 6 (Continue).

Let H be the *directed* graph obtained as follows: Each node v in G chooses exactly one node w with probability $b_{v,w}$ from its neighbors, while it selects no neighbors with probability $1 - \sum_{w \in \Delta_v} b_{v,w}$. Such a neighbor becomes its *incoming* edge.

We consider the following diffusion model: given a set of nodes A_0, the set $\bar{A}_t$ of active nodes at step t is given by the set of nodes that are reachable from A_0 (including A_0) in at most t steps.

Proof of Theorem 6 (Continue).

Let H be the *directed* graph obtained as follows: Each node v in G chooses exactly one node w with probability $b_{v,w}$ from its neighbors, while it selects no neighbors with probability $1 - \sum_{w \in \Delta_v} b_{v,w}$. Such a neighbor becomes its *incoming* edge.

We consider the following diffusion model: given a set of nodes A_0, the set $\bar{A}_t$ of active nodes at step t is given by the set of nodes that are reachable from A_0 (including A_0) in at most t steps.

Let $\bar{\sigma}_H(A)$ be the number of active nodes at the end of such a process. $\bar{\sigma}_{H(\cdot)}$ is submodular (set cardinality is submodular) and therefore $\bar{\sigma}(\cdot)$ is submodular (non-negative linear combination of submodular functions).

Proof of Theorem 6 (Continue).

Moreover:

$$Pr(v \in \bar{A}_{t+1} | v \notin \bar{A}_t) = \frac{\sum_{u \in \bar{A}_t \setminus \bar{A}_{t-1}} b_{v,u}}{1 - \sum_{u \in \bar{A}_{t-1}} b_{v,u}}.$$

By induction on t, we can therefore prove that

$$Pr(v \in \bar{A}_{t+1} | v \notin \bar{A}_t) = Pr(v \in A_{t+1} | v \notin A_t),$$

for all t. Hence, $\sigma(\cdot) = \bar{\sigma}(\cdot)$ which proves that $\sigma(\cdot)$ is submodular. $\square$

Influence Maximization

- An approximation to the influence maximization problem can be obtained by means of the greedy hill-climbing algorithm in both the IC and LT model.
- The most difficult part is to evaluate the marginal gain $f(S \cup \{v\} - f(S))$, which in practice is done by running simulations.
- This strategy or similar strategies are used at Facebook, Twitter, etc.

References I

[1] M. Mitzenmacher and E. Upfal.
Probability and Computing : Randomized Algorithms and Probabilistic Analysis.
Cambridge University Press, New York (NY), 2005. Section 1.3, page 9.

[2] Nemhauser, George L., Laurence A. Wolsey, and Marshall L. Fisher.
An analysis of approximations for maximizing submodular set functions.
Mathematical Programming 14.1 (1978): 265-294.

[3] Kempe, David, Jon Kleinberg, and Eva Tardos.
Maximizing the spread of influence through a social network.
Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2003.

[4] G. Nemhauser, L. Wolsey, M. Fisher.
An analysis of the approximations for maximizing submodular set functions.
Mathematical Programming, 14(1978), 265?294

References II

[5] Chierichetti, Flavio, Silvio Lattanzi, and Alessandro Panconesi.
Rumor spreading in social networks.
Theoretical Computer Science 412.24 (2011): 2602-2610.