Modelling influence propagation in social networks

Anna Szczepanek

Jagiellonian University, Institute of Mathematics

Based on Master’s thesis supervised by Marcin Mazur

tfml 2015
theoretical foundations of machine learning, Będlewo
Introduction

- influence propagation in social networks
- applications - propagation of ideas & behaviours, adoption of technology, viral marketing
- KKT - influence maximisation problem


D. Kempe, J. Kleinberg, E. Tardos. *Influential nodes in a diffusion model for social networks*. ICALP’05. Springer-Verlag, 2005
**Introduction**

**General Framework**

**Cascade Models**

**Threshold models**

**Generalised Models**

**Summary**

---

\[ G \] - network of individuals (directed graph)

\[ \downarrow \]

\[ S_0 \] - initial set of \( k \) individuals (\( k \) is a parameter)

\[ \downarrow \]

**random** propagation

\[ \downarrow \]

\[ \varphi(S_0) \] - final set of influenced (also: active) individuals

---

**Influence Maximisation Problem:**

find \( S_0 \) which maximises

*(the expected value of)*

the cardinality of \( \varphi(S_0) \)
(Ω, 2Ω, ℙ) - a fixed probability space, Ω is finite.

*Social network:* a directed graph $G = (V, E)$

- vertices $V = \{v_1, \ldots, v_N\}$
- edges $E \subset (V \times V) \setminus \{(v, v) \mid v \in V\}$

*Propagation:* A family of stochastic processes $\mathcal{P} := \{\mathcal{P}_S\}_{S \in 2^V}$, where $\mathcal{P}_S: \Omega \times \mathbb{N} \ni (\omega, i) \mapsto S_i(\omega) \in 2^V$ such that

1. $S_0(\omega) := S$
2. $v \in S_i(\omega) \setminus S_{i-1}(\omega) \Rightarrow$ there exists $u \in S_{i-1}(\omega)$ and $(u, v) \in E$
**Influence function**: expected final number of active vertices when starting from a given set

\[ \sigma : 2^V \rightarrow \mathbb{N}, \quad \sigma(S) = \mathbb{E}(\#S_T), \quad S_T \ - \text{final set} \]

**Influence maximisation problem** with parameter \( k \in \mathbb{N} \):
to find \( S^* \) such that \( \sigma(S^*) = \max\{\sigma(S) \mid S \subset V, \ #S = k\} \)

the influence function \( \sigma : 2^V \rightarrow \mathbb{N} \) is called

- **monotone** if \( \sigma(S) \leq \sigma(\bar{S}) \)
- **submodular** if \( \sigma(S \cup \{v\}) - \sigma(S) \geq \sigma(\bar{S} \cup \{v\}) - \sigma(\bar{S}) \)

for all sets \( S \subset \bar{S} \)
Theorem (Nemhauser, Wolsey, Fisher (1978); KKT (2003))

If $\sigma: 2^V \rightarrow \mathbb{N}$ is monotone and submodular then the set of vertices $S$ chosen by the greedy algorithm satisfies $\sigma(S) \geq (1 - 1/e)\sigma(S^*)$, where $S^*$ is the solution to the influence maximisation problem with parameter $k$. ($(1 - 1/e) \approx 63\%$)

By the **greedy algorithm** we mean:

1. $S := \emptyset$
2. for $i = 1$ to $k$
   . $v_i \leftarrow \arg\max_{v \in V \setminus S} (\sigma(S \cup \{v\}) - \sigma(S))$
   . $S \leftarrow S \cup \{v_i\}$
Independent Cascade Model

For \((u, v) \in E\) we have \(p_{(u, v)} \in (0, 1]\) and a random variable

\[X_{(u, v)} : \Omega \to \{0, 1\} \quad (was \ the \ activation \ attempt \ successful?)\]

\[
\mathbb{P}(X_{(u, v)} = 1) = \begin{cases} p_{(u, v)} & (u, v) \in E \\ 0, & (u, v) \notin E \end{cases}
\]

These random variable are assumed to be independent.

\[S_i(\omega) := S_{i-1}(\omega) \cup A_i(\omega)\]
\[A_0(\omega) := S\]
\[A_{i+1}(\omega) := \{v \in V \setminus S_i(\omega) \mid \exists u \in V \ X_{(u, v)}(\omega) = 1 \land u \in A_i(\omega)\}\]
Theorem (KKT’03)

The influence maximisation problem in ICM is NP-hard.

Theorem (KKT’03)

The influence function in ICM is monotone and submodular.

Generalised Cascade Model (GCM)

\[ X_v := (X_{(u_i,v)} \mid u_i \in V, \ i = 1, \ldots, N) \] are independent

Parameters: \[ p_v(u, S) := \mathbb{P}(X_{(u,v)} = 1 \mid \forall w \in S \ X_{(w,v)} = 0) \]
**Increasing Cascade Model** (IncrCM).

$S \subset \tilde{S} \implies p_v(u, S) \leq p_v(u, \tilde{S})$

**Decreasing Cascade Model** (DCM) /appeared already in KKT/

$S \subset \tilde{S} \implies p_v(u, S) \geq p_v(u, \tilde{S})$
Threshold models

[Diagram of a network with nodes and arrows indicating influence propagation]
Threshold models
Threshold models
Threshold models

Anna Szczepanek

Modelling influence propagation in social networks
Threshold models
Threshold models
Threshold models
**Linear Threshold Model**

For every edge in $G = (V, E)$ we have $b_{(u,v)} \in (0, 1]$ such that for all $v \in V$

$$\sum_{u: (u,v) \in E} b_{(u,v)} \leq 1 \text{ (normalisation)}$$

**Thresholds:** $\theta_v : \Omega \to [0, 1]$, independent, uniformly distributed

**Active sets:**

$$S_i(\omega) := S \cup \{v \in V | \sum_{u \in S_{i-1}(\omega)} b_{(u,v)} \geq \theta_v(\omega)\}$$

*linear accumulation of influence*
Linear Threshold Model

For every edge in $G = (V, E)$ we have $b_{(u,v)} \in (0, 1]$ such that for all $v \in V$

$$\sum_{u: (u,v) \in E} b_{(u,v)} \leq 1 \text{ (normalisation)}$$

Theorem (KKT’03)

*The influence maximisation problem in LTM is NP-hard.*

Theorem (KKT’03)

*The influence function $\sigma$ in LTM is monotone and submodular.*
Non-normalised Linear Threshold Model

For every edge in $G = (V, E)$ we have $b_{(u,v)} \in (0, 1]$ such that for all $v \in V$

$$\sum_{u: (u,v) \in E} b_{(u,v)} \leq 1 \text{ (normalisation)}$$
Non-normalised Linear Threshold Model

For every edge in $G = (V, E)$ we have $b_{(u, v)} \in (0, 1]$
Non-normalised Linear Threshold Model

For every edge in $G = (V, E)$ we have $b_{(u,v)} \in (0, 1]$

- uniform rescaling to LTM may lead to a wrong solution
Generalised Threshold Model

We consider for each $v$ the *activation function*

$$f_v : 2^N_v \to (0, 1]$$

which is assumed to be monotone and to satisfy $f_v(\emptyset) = 0$.

**Active sets:**

$$S_i(\omega) := S \cup \{v \in V \mid f_v(S_{i-1}(\omega)) \geq \theta_v(\omega)\}$$

In LTM:

$$f_v(S) = \sum_{u \in S} b(u, v)$$

In nLTM:

$$f_v(S) = \min\left(\sum_{u \in S} b(u, v), 1\right)$$
Locally Submodular Threshold Model

All the activation functions $f_v$ are required to be submodular.

The submodularity of $\sigma$ in this model was conjectured by KKT (2003) and proved by Mossel & Roch (2010).

In nLTM we have $f_v(S) = \min(1, \sum_{u \in S} b_{u,v})$ - submodular!
Generalised Models: GCM & GTM (KKT)
Generalised Models: GCM & GTM (KKT)
Generalised Models: GCM & GTM (KKT)
Results about DCM & IncrCM

DCM  the influence function is submodular (delayed propagation processes in KKT’05)
DCM  it is a special case of LocSTM

IncrCM  the influence function is not submodular
IncrCM  it is not a special case of LocSTM

necessary condition for LocSTM:

\[ f_v(S) \leq \sum_{u \in S} f_v(u) \]

IncrCM  nLTM is a special case of IncrCM & it is submodular
The formalised framework for models of influence propagation:
- unifies cascade and threshold models & clarifies their structure
- reveals new connections & leads to new models