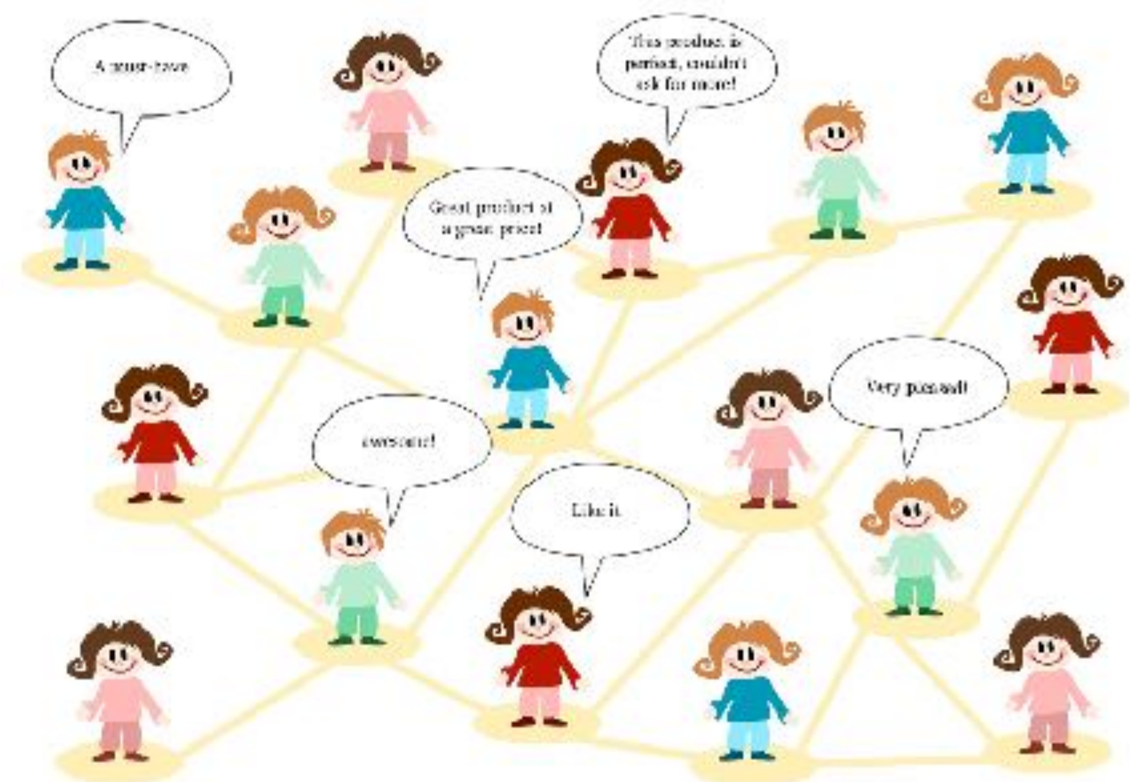
Model-Independent Online Learning for Influence Maximization

Sharan Vaswani[1], Branislav Kveton[2], Zheng Wen[2], Mohammad Ghavamzadeh[3],
Laks Lakshmanan[1], Mark Schmidt[1]
[1] University of British Columbia [2] Adobe Research [3] Deepmind

Influence Maximization

Underlying principle: Influence propagates through ‘word of mouth’ in a social network



Idea: Give discounts to ‘influential’ users who will trigger off word-of-mouth epidemics

Aim: Find a subset of users (‘seed set’) who will influence maximum people to become aware of a product

Influence Maximization

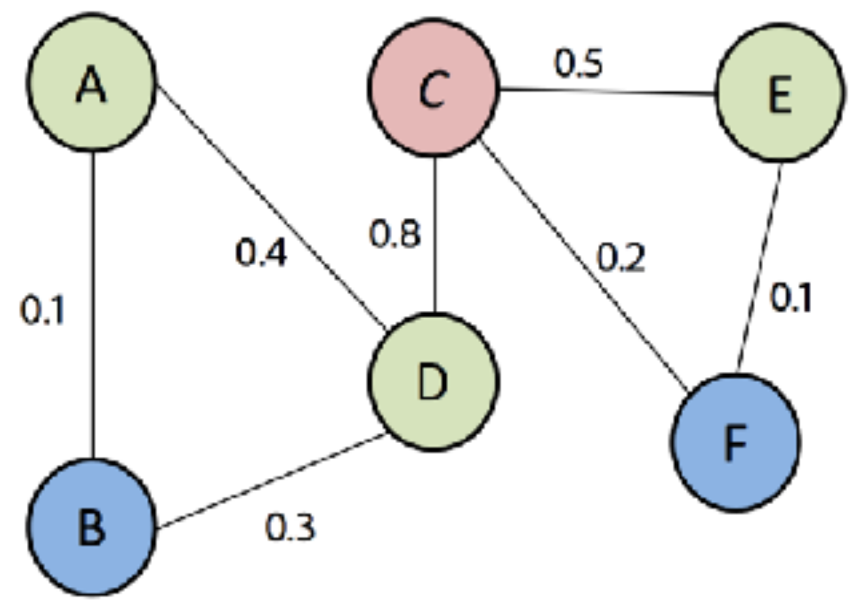
Input: $(\mathcal{G}, \mathcal{C}, \mathcal{D})$

Stochastic diffusion Model

Weighted Graph

Set of feasible seed sets

E.g: $\mathcal{C} \subseteq \{\mathcal{S} \subseteq \mathcal{V} : |\mathcal{S}| \leq K\}$



Objective: Find $\mathcal{S}^* \in \arg\max_{\mathcal{S} \in \mathcal{C}} F(\mathcal{S}) = \sum_{v \in \mathcal{V}} F(\mathcal{S}, v)$

Expected number of nodes influenced by S

Pr (S influences target node v)

Key Property: For common diffusion models, $F(\mathcal{S})$ is submodular in S

Influence Maximization

Challenges to using IM in practice:

- **Challenge 1:** IM is not robust to the choice of the diffusion model [1] nor its model parameters [2].

- **Challenge 2:** Learning model parameters requires considerable data, often unavailable to a new marketer.

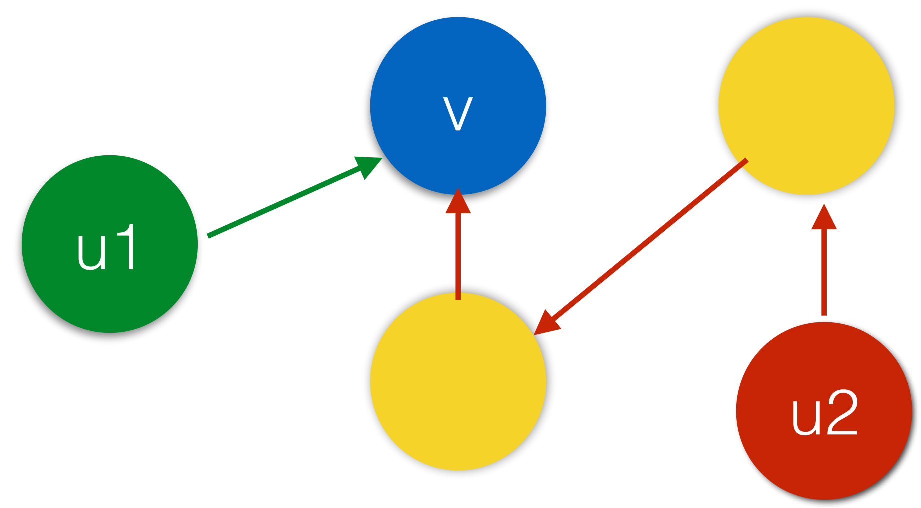
[1] Du, Nan, et al. "Influence function learning in information diffusion networks." ICML, 2014
[2] Goyal, Amit, et al. "Learning influence probabilities in social networks." WSDM, 2010

Model Independent Formulation

Assumption 1: F(S) is monotonic in S.

Key Idea: Parametrize the problem in terms of pairwise reachability probabilities $p^*_{u,v} = F(\{u\}, v)$ ← Pr(u influences v under a diffusion model)

Surrogate Objective: Find $\widetilde{S} \in \arg\max_{\mathcal{S} \in \mathcal{C}} \left[\sum_{v \in \mathcal{V}} \left(\max_{u \in \mathcal{S}} p^*_{u,v} \right) \right]$



Advantages:

- Common parametrization for all progressive models.
- Guaranteed approximation.
- Surrogate objective $f(\mathcal{S}, p^*)$ is submodular irrespective of the diffusion model.

Influence Maximization

Challenges to using IM in practice:

- **Challenge 1:** IM is not robust to the choice of the diffusion model [1] nor to the model parameters [2].
- **Challenge 2:** Learning model parameters requires considerable data, often unavailable to a new marketer.

Online Influence Maximization

Setting: New marketer who has no past data to learn the reachability probabilities

Idea: Perform IM while simultaneously learning $p^*_{u,v}$ through trial and error across multiple rounds.

Basic Protocol:

for $t = 1$ **to** T **do**

submodular optimization subroutine

probability estimates

- Choose $\mathcal{S}_t \leftarrow \texttt{ORACLE}(\mathcal{G}, \mathcal{C}, \overline{p})$

- Diffusion occurs according to an underlying diffusion model.

- Observe semi-bandit feedback.

  for $u \in \mathcal{S}_t$ **do**

  size n binary vector. each entry = 1 iff that node is influenced by the seed u

  Get pairwise influence feedback $\boldsymbol{y}_{u,t}$

- Update parameter estimates $\overline{p}_{u,v}$

Online Influence Maximization

Challenge 1: Learn n^2 reachability probabilities

Assumption 2. *For all $u, v \in \mathcal{V}$, $p^*_{u,v}$ can be "well approximated" by the inner product of $\boldsymbol{\theta}^*_u$ and $\boldsymbol{x}_v$, i.e.,*

$$p^*_{u,v} \approx \langle \boldsymbol{\theta}^*_u, \boldsymbol{x}_v \rangle \triangleq \boldsymbol{x}_v^\top \boldsymbol{\theta}^*_u$$

d dimensional feature describing a target node
(Eigenbasis features, node2vec [3])

Vector to be learnt for every source node.

Advantages:

- Reduces the number of parameters from $O(n^2)$ to $O(dn)$.
- In each round, mean estimates of $\overline{p}_{u,v}$ can be updated by solving K regression problems.

[3] Grover, Aditya, et al. "node2vec: Scalable feature learning for networks." KDD, 2016.

Online Influence Maximization

Challenge 2: Trade off exploration and exploitation

Basic Idea:
Use the Upper Confidence Bound algorithm i.e. use overestimate (mean + variance) of reachability probabilities as input to the oracle.

Computational Complexity:

UCB computation oracle computation

Per-round time: $\underbrace{O(Knd^2)}_{\text{UCB computation}} + \underbrace{O(Kn)}_{\text{oracle computation}}$

Online Influence Maximization

Performance metric:

$$R^{\kappa}(T) = T \cdot F(S^*) - \frac{1}{\kappa}\mathbb{E}\left[\sum_{t=1}^{T} F(S_t)\right]$$

Regret after T rounds

Optimal seed set in hindsight

To account for the approximation in ORACLE

Selected seed set

Regret Bound:

standard combinatorial bandit dependence

near optimal dependence

$$R^{\rho\alpha}(T) = \tilde{O}(n^2 d\sqrt{KT}/(\alpha\rho))$$

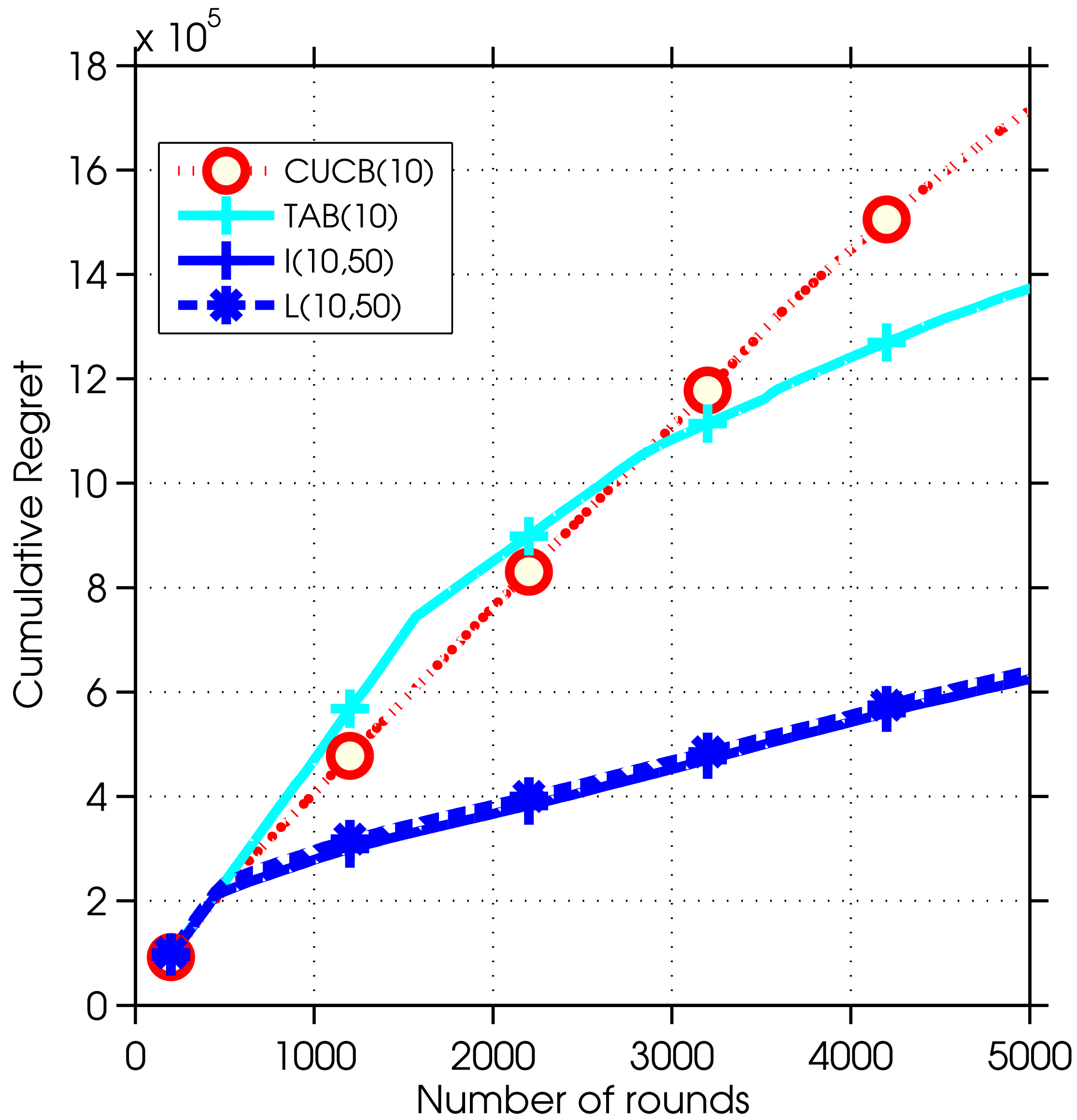
Surrogate approximation factor

best dependence on network size

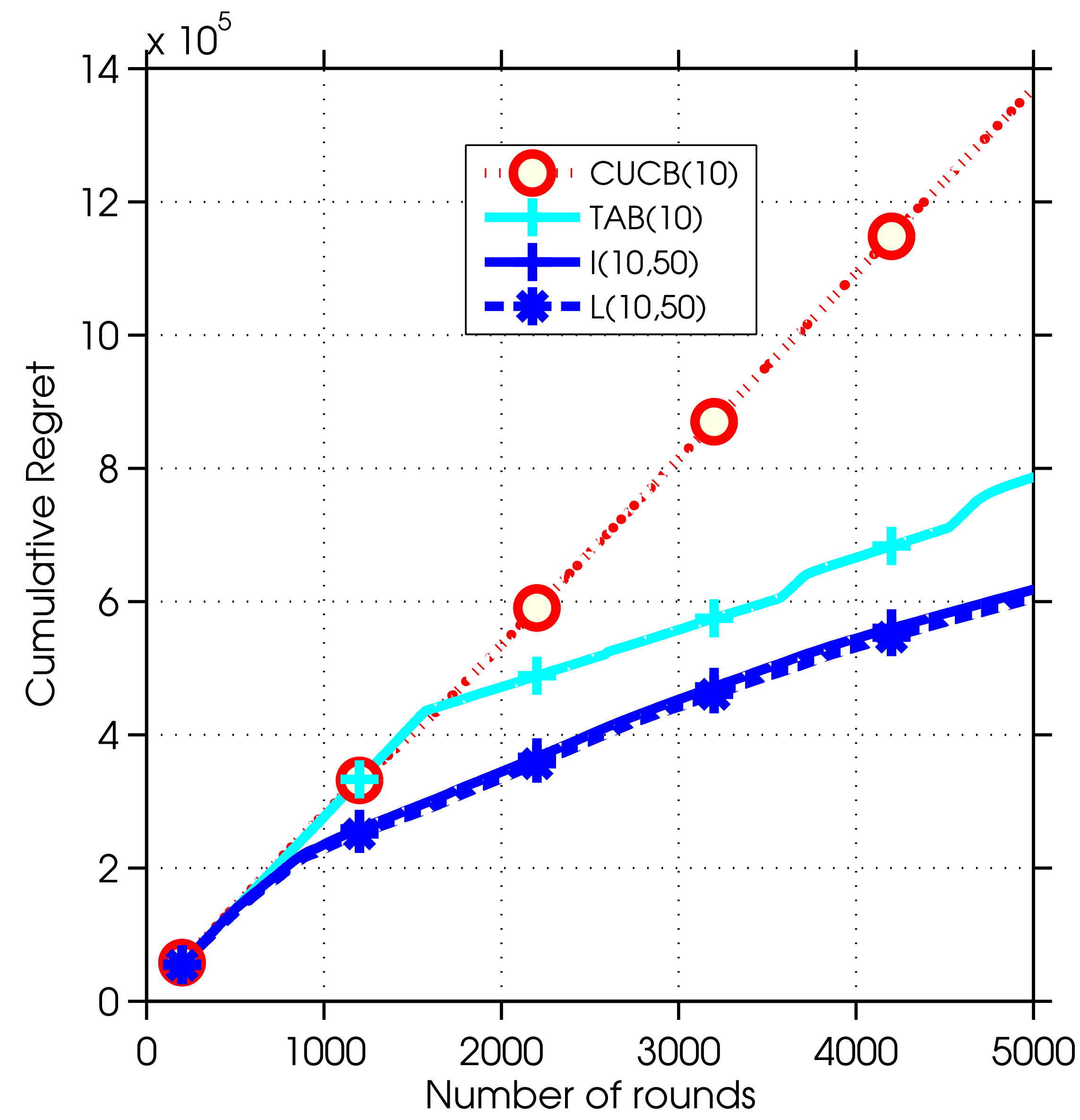
ORACLE approximation factor

standard linear bandit dependence

Experiments on Facebook dataset



IC model



LT model

Conclusion

Contributions:

- Developed a model-independent parametrization for IM and proposed a surrogate objective function.

- Proposed and analyzed a UCB based algorithm for model-independent online IM.

Future Work:

- Extend the framework to different feedback models and bandit algorithms.

- Generalization across source nodes for better statistical efficiency.