A Heuristic Approach for Viral Marketing Cost Optimization in Social Networks

Ashis Talukder, and Choong Seon Hong
Department of Computer Science and Engineering, Kyung Hee University, South Korea
Email: {ashis, cshong}@khu.ac.kr

Abstract

Nowadays, social networks have become a very powerful means of Viral Marketing (VM), and the Influence Maximization (IM) is such a viral marketing approach which determines the viral marketing profit. The profit is defined by the maximum number of nodes that can be influenced by the seed nodes in the network. However, most of the existing IM models assume that the seed users are initially activated and thus, do not focus on optimizing the Viral Marketing Cost (VMC) which is defined by the minimum number of nodes required to activated the seed users. Although some Reverse influence Maximization (RIM) models are available in the literature, their accuracy is not satisfactory. Thus, in this paper, we propose a Heuristic RIM (H-RIM) model in which Independent Cascade (IC), Linear Threshold (LT), and Greedy models are applied together to minimize the VM cost. The proposed H-RIM model handles RIM-challenges efficiently as well as returns optimized VM cost. The simulation on real network datasets shows that the proposed model outperforms the existing models.

1. Introduction

Nowadays, social networks are considered to be the most robust platform for Viral Marketing (VM), and the Influence Maximization (IM) is such a viral marketing tool that estimates influential nodes which can maximize the influence diffusion in the network [1]. The influence is measured by the maximum number of nodes that can be activated by the seed users when they are assumed to be activated initially. However, most of the IM models focus on profit and thus, ignore the seeding cost or Viral marketing cost (VMC). Thereafter, a Reverse Influence Maximization (RIM) technique is emerged to minimize the VM cost [2], [3]. However, the existing RIM models fail to exhibit the expected accuracy to minimize the VM cost and meet the RIM challenges [4].

Therefore, in this paper, to address the VMC Minimization problem, we propose a Heuristic RIM (H-RIM) model which jointly employs the Independent Cascade (IC) model, the Linear Threshold (LT) model [1], and a greedy method. The key contributions of this paper are stated below.

1) We employ the traditional IC model in reverse order for node activation in the H-RIM model. The use of the IC and greedy models resolve most of the RIM challenges.
2) We introduce a cost minimization heuristic which is used in the influence weight estimation and is applied with the greedy optimization to minimize the VM cost.
3) The simulation, done on real datasets, shows that the proposed model outperforms the existing models.

2. Existing Study

The significant development in the Influence Maximization (IM) research is contributed by Kempe *et al.* [1] who formulate the Linear Threshold (LT) and Independent Cascade (IC) models. After that, many profit maximization models are proposed using the IM technique. A profit maximization model by increasing the product adoption is proposed in [5]. Lu *et al.* [6] propose a product adaptation model as well; however, they consider the product price as a criterion to adopt a product. While maximizing the profit, Zhu *et al.* [7] observe that the profit and influence cannot be maximized together whereas, Goyal *et al.* [8] propose a recommendation-based profit maximization model. However, none of the above studies addresses the VM cost estimation problem.

Then, the Reverse Influence Maximization (RIM) models which determine the minimum number of nodes needed to activate seed nodes, are proposed in [2] and [3]. However, the accuracy of the existing RIM models is not satisfactory and thus, in this paper, we propose a Heuristic RIM (H-RIM) model for VMC Minimization. The proposed model meets RIM challenges efficiently and estimates the optimized VM cost as well.

3. System Model

Here, we maximize the viral marketing profit in social networks by minimizing the cost. To formulate the VMC minimization problem, we take a social network, $G(V,E)$ having a vertex set V, and an edge set E. We denote the in-neighbor and out-neighbor sets of any node v by D_v^{in} and D_v^{out}, respectively. For a given seed set S, the viral marketing cost, denoted as $\lambda(S)$, is defined by the number of users requires to activate the seed nodes. The objective of the VMC minimization problem is to minimize the $\lambda(S)$.

Definition 1 (The VMC minimization Problem)**.** Given a social network $G(V,E)$ and a seed set S of size k, the aim of the VMC minimization problem is to minimize the viral marketing cost $\lambda(S)$, defined by the minimum number of nodes that must be activated in order to activate all the seed nodes in S. □

4. The Proposed Heuristic RIM (H-RIM) Model

Here, we propose the Heuristic Reverse Influence Maximization (H-RIM) Model to estimate the optimal VM cost. First, the partial VM cost $\lambda(u)$ for all $u \in D_v^{in}$ is calculated by using the IC model used in a retrograde manner and then, the partial cost is optimized by the greedy heuristic technique and the LT model to estimate $\lambda(v)$.

Algorithm 1: The H-RIM Algorithm

Input: $G(V, E), S$
Result: $\lambda(S)$

$\Lambda(S) := \emptyset$
for $v \in S$ **do**
  $\Lambda(v) := \emptyset, A_{new} := \{u\}$;
  while $u \in A_{new}$ **do**
    $A_{target} := D_u^{in}, A_{current} := \emptyset, \Lambda(u) := \{u\}$;
    while $w \in A_{target}$ **do**
      if *w is activated with probability p* **then**
        $A_{current} := A_{current} \cup \{w\}$; /* Node activated by RIC model */
      end
    end
    $A_{new} := A_{new} - \Lambda(u) - \Lambda(S)$; /* Already activated node is excluded */
    $\Lambda(u) := \Lambda(u) \cup A_{new}$;
  end
  A_v := A minimum set of $u \in D_v^{in}$ with max $\tilde{w}_{uv}$ to activate v; /* Greedy optimization */
  $\Lambda(v) := \cup_{y \in A_v} \Lambda(y)$;
  $\Lambda(S) := \Lambda(S) \cup \Lambda(v)$; /* The VM cost set */
end
$\lambda(S) := \big|\Lambda(S)\big|$; /* Final VM cost */
return $\lambda(S)$;

A. Partial VM cost estimation

The traditional IC model is applied in reverse order, and the node activation is performed to estimate $\lambda(u)$, as depicted in Figure 1 (the left block). We calculate $\lambda(u)$ hop by hop and initially, we assume that only u is the only activated node and determine which nodes are necessary to activate u by IC model. Thus, the newly and total activated nodes are listed as,

$$A_{new} = \Lambda(u) = \{u\} \tag{1}$$

For each $u \in A_{new}$, the target in-neighbor set, $A_{target}(u)$ is calculated as:

$$A_{target}(u) = D_u^{in} \tag{2}$$

Then, each $w \in A_{target}(u)$ is given a single chance to activate u with a probability p. Here, the value of p is selected as $p \in \{0.01, 0.1\}$ [1], [9]. If the node w activates the node u then w is included in $A_{current}$. Then, at the end of hop, we update the newly and total activated nodes for the next hop as:

$$A_{new} = A_{current} - \Lambda(u) \tag{3}$$

$$\Lambda(u) = \Lambda(u) \cup A_{current} \tag{4}$$

Here, we exclude the nodes which are already activated in any previous hop as shown in (3). The same process is repeated for the next in-neighbor hops of u for each of the newly activated node w and so on. The IC process terminates when there is no newly activated node at any hop ($A_{new} = \emptyset$).

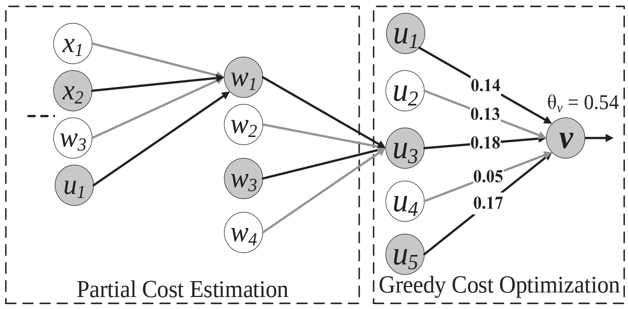


Figure 1: The working strategy of the H-RIM model.

B. The Optimization Model

Now, to compute $\Lambda(v)$, we will have to select a set A_v of minimum number of $u \in D_v^{in}$ to activate v as shown in Figure 1 (the right block) by the greedy method.

We use the LT model in reverse order in which the influence weight is estimated by a combination of the heuristic weight and the social influence weight as,

$$hw_{uv} = \alpha h_{uv} + (1 - \alpha) w_{uv} \tag{5}$$

where, $\alpha \in (0, 1)$, is a constant, and w_{uv} is the social influence weight computed by the degree centrality method [1].

We contribute a heuristic influence h_{uv}, which is the normalized value of inverse proportional to the estimated $|\Lambda(u)|$. *The lower cost $|\Lambda(u)|$, of u indicates the higher value of h_{uv}.* The threshold θ_v is selected by the majority voting model (for example, in Figure 1, $\theta_v = 3$) [10]. Finally, we calculate the optimized VM cost as,

$$\Lambda(v) = \underset{u \in A_v}{\cup} \Lambda(u) \tag{6}$$

$$\lambda(S) = \big|\Lambda(S)\big| = \left|\underset{v \in S}{\cup} \Lambda(v)\right| \tag{7}$$

The H-RIM model is stated in the Algorithm 1 and the running time of the algorithm is given by,

$$C \leq k(d(d + d)) \approx O(kd^2) \tag{8}$$

Theorem 1. *The VMC minimization problem is NP-Hard.*

Proof. The greedy technique used in the H-RIM model to solve the VMC minimization problem is the Knapsack technique, and at each time, the model selects an in-neighbor u with maximum influence hw_{uv}. Moreover, the Knapsack problem is a well-known NP-Hard [11] problem and hence, our VMC minimization problem under the HM model is also NP-Hard. □

5. Performance Evaluation

We evaluate the performance of the proposed H-RIM model by using datasets of two important social networks.

A. Data Collection

We collect real datasets of the Epinions[1] and the Facebook[2] network from the SNAP dataset collections [12], as shown in Table I.

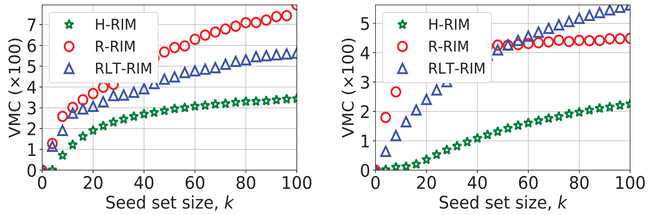
[1] https://snap.stanford.edu/data/soc-Epinions1.html
[2] https://snap.stanford.edu/data/egonets-Facebook.html

Table I: Datasets

Network Name	Nodes	Edges
Epinions	$75,879$	$508,837$
Facebook	$4,039$	$88,234$

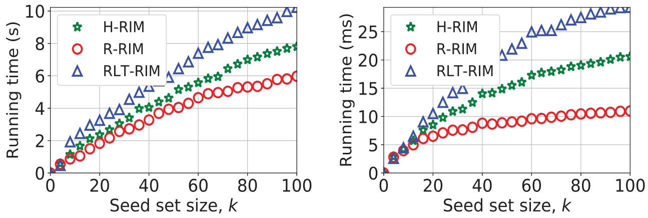
B. Simulation Setup

We employ the *Monte Carlo (MC)* simulation by using Python codes. The seed set S is generated randomly and the probability is chosen as $p \in \{0.1, 0.01\}$ [1]. We select θ_v by the Heuristic Individual (HI) model [13] and $\alpha = 0.5$ [14]. We compare the results with that of existing R-RIM and RLT-RIM models [2], [3].



(a) Epinions dataset. (b) Facebook dataset.

Figure 2: The estimated VM cost for different values of k.



(a) Epinions dataset. (b) Facebook dataset.

Figure 3: The running time for different values of k.

C. Viral Marketing Cost

The estimated viral marketing cost for different datasets is presented in Figure 2. The figure shows that the proposed H-RIM model has VM cost $2-3$ times smaller than that of the existing R-RIM and RLT-RIM algorithms, for both the datasets. This optimized result is due to the use of the heuristic greedy optimization technique.

D. Running Time

Figure 3 depicts the running time requirement of the proposed algorithm along with the existing R-RIM and RLT-RIM models for both the datasets. The R-RIM requires the least time, and the RLT-RIM consumes the highest time. On the other hand, the running time of the proposed H-RIM model lies between that of two existing models.

E. Handling RIM Challenges

The proposed H-RIM model shows better accuracy by handling the RIM challenges efficiently as well as providing the minimized VM cost. The use of IC model in the node activation process addresses the issue of Basic Network Components (BNC) and the setting terminating condition [2]. The greedy model takes care of the NP-Hardness of the problem. The use of the majority voting model to assign the node threshold resolves the insufficient influence [2].

6. Conclusion

In this paper, we propose a Heuristic Reverse Influence Maximization (H-RIM) model to address the Viral Marketing Cost (VMC) minimization problem. We use the Independent cascade (IC) model in reverse order for node activation process and the greedy heuristic model along with the LT technique to optimize the VM cost. We evaluate the performance of the proposed model by using the datasets of two widely known social networks, and the empirical result shows that the proposed model outperforms the existing models.

Acknowledgment

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF-2017R1A2A2A05000995).
***Professor Dr. CS Hong** is the corresponding author.

References

[1] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 137–146.

[2] A. Talukder, M. G. R. Alam, A. K. Bairagi, S. F. Abedin, M. A. Layek, H. T. Nguyen, and C. S. Hong, "A cost optimized influence maximization in social networks," in *2017 IEEE The 19th Asia-Pacific Network Operations and Management Symposium (APNOMS 2017)*. IEEE, 2017.

[3] A. Talukder, M. G. R. Alam, H. T. Nguyen, and C. S. Hong, "A cost optimized reverse influence maximization in social networks," in *2018 IEEE/IFIP Network Operations and Management Symposium (NOMS 2018)*. IEEE, 2018.

[4] A. Talukder and C. S. Hong, "Active reverse path based reverse influence maximization in social networks," in *Korea Software Congress (KSC 2017)*. KIISE, 2017, pp. 1203 – 1205.

[5] S. Bhagat, A. Goyal, and L. V. Lakshmanan, "Maximizing product adoption in social networks," in *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 2012, pp. 603–612.

[6] W. Lu and L. V. Lakshmanan, "Profit maximization over social networks," in *Data Mining (ICDM), 2012 IEEE 12th International Conference on*. IEEE, 2012, pp. 479–488.

[7] Y. Zhu, Z. Lu, Y. Bi, W. Wu, Y. Jiang, and D. Li, "Influence and profit: Two sides of the coin," in *2013 IEEE 13th International Conference on Data Mining*. IEEE, 2013, pp. 1301–1306.

[8] A. Goyal and L. V. Lakshmanan, "Recmax: Exploiting recommender systems for fun and profit," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 1294–1302.

[9] A. Talukder and C. S. Hong, "Epidemiological reverse influence maximization in social networks with negative influencing," in *Korean Computer Congress (KCC 2018)*. KIISE, 2018, pp. 1277–1279.

[10] A. Talukder, A. K. Bairagi, D. H. Kim, and C. S. Hong, "Reverse path activation-based reverse influence maximization in social networks," *Journal of The Korean Institute of Information Scientists and Engineers (JOK)*, vol. 45, no. 11, pp. 1203–1209, 2018.

[11] E. Horowitz and S. Sahni, *Fundamentals of computer algorithms*. Computer Science Press, 1978.

[12] J. Leskovec and A. Krevl, "SNAP Datasets: Stanford large network dataset collection," http://snap.stanford.edu/data, jun 2014.

[13] A. Talukder, M. G. R. Alam, A. K. Bairagi, S. F. Abedin, H. T. Nguyen, and C. S. Hong, "Threshold estimation models for influence maximization in social network," in *Korean Institute of Information Scientists and Engineers (KIISE)*. KIISE, 2016, pp. 888–890.

[14] A. Talukder, R. Kamal, A. K. Bairagi, M. G. R. Alam, S. F. Abedin, M. A. Layek, H. T. Nguyen, and C. S. Hong, "Rumors in the social network: Finding the offenders using influence maximization," in *Korean Computer Congress (KCC)*. KIISE, 2015, pp. 1214–1216.