Applied Soft Computing xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Sequential seeding to optimize influence diffusion in a social network

Yaodong Ni*

School of Information Technology and Management, University of International Business and Economics, Beijing 100029, China

ARTICLE INFO

Article history:
Received 4 December 2015
Received in revised form 19 February 2016
Accepted 19 April 2016
Available online xxx

Keywords:
Social networks
Complete influence time
Markov decision process
Online algorithm
Modified greedy algorithm

ABSTRACT

The problem of node seeding for optimizing influence diffusion in a social network can be applied in many fields, and thus has drawn much attention. In real life, because of a variety of reasons, decision maker needs to make a sequence of decisions about how to select the seeded nodes. In this paper, we study the problem of sequentially seeding nodes in a social network such that the complete influence time is minimized. We formulate a Markov decision process to describe the problem and embed a modified greedy search method into an online algorithm to solve the Markov decision process. Numerical experiments are performed to show the effectiveness of the proposed online algorithm.

1. Introduction

In recent years, with the extensive adoption of social network services, the massive records of human interactions make it possible to analyze and optimize influence diffusion through social networks. The problem of optimizing influence diffusion in a social network concentrates on how to select a small number of individuals as the initial influence adopters to trigger a cascade such that the influence diffusion in the social network is optimized. This problem has drawn much attention, as it can be applied to optimize the spread of information, new idea or innovation through social networks. One of the most popular applications is viral marketing [6,9,12]. For a firm who needs to promote a new product, viral marketing is an effective strategy that can help to save promotion costs. Instead of promoting new product to all potential consumers, a firm can seed a small number of consumers who would recommend the product to their peers. Due to the word-of-mouth effect, the product can be widely accepted.

Recently, the literature on influence diffusion optimization has been growing. A number of models for influence diffusion have been proposed and studied [16,25,30,32,28]. There have been a great number of papers that focus on algorithm design for influence diffusion optimization [19,13,34,4,21,33,18,23,1]. Different objectives for optimizing influence diffusion have been proposed, such as maximizing the number of the ultimately influenced nodes [16,17], maximizing the expected lift in profit [9,29], minimizing the size of the target set of nodes that ensures complete influence [7], minimizing the number of seeded nodes that ensure at least a given number of individuals are eventually influenced [22], minimizing the complete influence time [25], to name a few. Besides, some researchers have investigated the problem of optimizing influence diffusion under the assumption that the budget for seeding individuals is limited [26,24,27].

In the works mentioned above, the problem of optimizing influence diffusion in social networks is studied from various aspects. However, very few papers focus on making sequential decisions for optimizing influence diffusion. In reality, as situation evolves and varies, we usually have to make sequential decisions. We take the application of viral marketing as an example. A firm usually needs a considerable amount of capital to seed the initial set of individuals for promoting a new product though a social network. However, more often than never, the capital may not be ready at the same time. The capital is usually separated into several parts that are available at different time points. In this view, the target set of nodes are sequentially constructed, and a sequence of decisions on seeding nodes need to be made. Even through the capital can be available at the very beginning, the management of the firm still has an incentive to seed sequentially. In the sense of optimizing the mathematical expectation of objective function, the best strategy for optimizing influence diffusion is known to be seeding as many nodes as possible at the very beginning. However, for each time of seeding under this strategy, the specific process of influence diffusion can be unpredictable, and the ultimate situation of influence diffusion in the social network may be far from optimality. In view of this, the management of the firm may prefer to seed

* Correspondence to: Room 1309, Boxue Building, University of International Business and Economics, Beijing 100029, China.
E-mail address: ni@uibe.edu.cn

http://dx.doi.org/10.1016/j.asoc.2016.04.025
1568-4946/

nodes sequentially, so that observation can be made before each time of seeding, based on which a reasonable further decision can be made. In summary, it is significant to study sequential seeding for optimizing influence diffusion.

In this paper, we study the problem of sequentially seeding nodes such that the influence diffusion in a social network is optimized. We focus on the influence diffusion optimization problem with the objective of minimizing the complete influence time [25]. For the cases where the influence propagated is dominant compared to existing counterparts, it is believed the influence will be ultimately adopted by all members in the social network. A typical instance is the replacement of Walkman by MP3 player. Compared with the former, the later is superior in almost every aspect, and thus it has been adopted by almost everyone nowadays. For such a case, a decision maker who attempts to optimize the influence is more concerned with shortening the period of time it takes to make every individual in the social network adopt the influence, which is defined as the complete influence time. In this paper, in order to characterize the process of influence diffusion in this case, we employ the incremental chance model proposed in [25]. We formulate the problem of sequential seeding for minimizing the complete influence time as a Markov decision process. Due to the tremendous numbers of states and actions in the formulated Markov decision process, traditional dynamic programming methods are not applicable. We design an online algorithm in which the modified greedy algorithm developed in [25] is embedded, and perform numerical experiments to show that the algorithm is effective.

As we mentioned, the existing papers on sequential seeding are very few. The research trend that is closest to our work is on the approach named adaptive seeding [31,15,2]. The motivation of adaptive seeding is to deal with the inapplicability of existing heuristic methods to the real applications where influential nodes are difficult to get access to. This approach divides the decision of seeding nodes into two sequential sub-decisions. In the first stage, the decision maker uses a part of budget to select a number of nodes and make them invite their neighbors to become the nodes that are accessible. In the second stage, the decision maker uses the remaining budget to seed a number of nodes who are in the larger accessible set of nodes. There are several differences between this research trend and our work. First, there are only two stages of seeding in the framework of adaptive seeding; while in our work the number of stages of decision making can be more than two. Second, the adaptive seeding approach assumes that not all nodes in the social network are accessible; while such an assumption is not made in our paper. Last, in the framework of adaptive seeding, the purpose of the former seeding is to enlarge the accessible set of nodes, and the later seeding is more significant; while in our work, later seeding stage is not more important than the former stages.

The remainder of this paper is organized as follows. Section 2 introduces the preliminary of social networks and incremental chance model. In Section 3, we present the problem of sequential seeding and formulate it as a Markov decision process. We propose an online algorithm based on a modified greedy algorithm in Section 4. Numerical experiments are performed in Section 5 to show the effectiveness of the proposed algorithm. Finally, conclusions are drawn in Section 6.

2. Social network and incremental chance model

A social network is a graph model that represents a group of individuals and their relationships. Mathematically, a social network is denoted by an indirected graph $G = (N, E, W)$, where N is the set of nodes, E is the set of edges and W is the weight function. Each node in N represents an individual. A pair of individuals is connected by a single edge in E, if they are related to each other. A positive value $W(i, j)$ denotes the weight on each edge (i, j) that quantifies the relationship between node i and node j. For any node i, node j is called a neighbor of i if there exists an edge (i, j) in E; and the set of neighbors of i is denoted by $N(i)$.

A number of models for influence diffusion have been proposed, such as the voter model [8,14], the majority rule model [11], the majority–vote model [20], the unanimity rule model [10], and so on. These models describe the symmetric dynamics that any state of an individual can be switched to any other state. Recently, more realistic cases were considered, where the process of influence diffusion is usually not symmetric but progressive. By progressive influence diffusion, we mean that there exists at least one state such that it will not switch to other states after it has been reached by any individual. A variety of papers on progressive models can be found [7,9,16,17,29], and typical ones include the linear threshold model [16], the independent cascade model [12], and the incremental chance model [25]. More often than never, a new product or a new idea that is dominant compared with the existing counterparts will be finally adopted by the whole population. For example, almost everyone today uses an MP3 player rather than a cassette tape Walkman. The incremental chance model is a natural mechanism for characterizing the influence diffusion in situations where the influence is dominant [25], which guarantees that all individuals in the social network are ultimately influenced, namely, complete influence. In this paper, we adopt the incremental chance model as the basic mechanism of the diffusion of influence.

In the incremental chance model, a node is either in active state representing that she has been influenced or in inactive state representing that she has not. For example, an active (inactive) node can denote an individual who has (has not) adopted an innovation. The incremental chance model characterizes the progressive dynamics that inactive nodes can turn active but the opposite will never happen. At each time step t, whether a node is influenced depends on a probability distribution, and the probability that an inactive node j turns active is given as

$$p_t^j = \frac{\sum_{i \in N_t^a(j)} W(i, j)}{\sum_{i \in N(j)} W(i, j)},$$

where $N_t^a(j)$ is the set of j's active neighbors at time t. The probability p_t^j is called the influence chance of node j at time t. The influence chance p_t^j for an active node j at any time t is defined as 1. We can find that the influence chance is nondecreasing and ultimately reaches 1. It has been shown that, under this model, the complete influence happens with probability 1, given that the initial target set of seeded nodes is non-empty [25].

It is not difficult to find that, under the incremental chance model, the complete influence happens only if each node outside the target set is connected by a path starting from at least one node in the target set. In this paper, the social network G is assumed to be connected, that is, there exists a path connecting any pair of nodes in G. This assumption is realistic and natural, because in real life nobody is absolutely isolated.

3. Sequential node targeting

Under the incremental chance model, the complete influence is always achieved. For this case, it is natural to study how to select and seed a target set of nodes to trigger the influence diffusion such that the complete influence time is minimized. For example, in viral marketing, given the limited capital, a firm needs to decide whom a new product should be promoted to such that the product will be widely adopted as soon as possible. This problem has been studied assuming that the volume of the target set is fixed, which