A linear threshold-hurdle model for product adoption prediction incorporating social network effects

Feng Zhou [a,b], Jianxin (Roger) Jiao [a], Baiying Lei [c,*]

[a] The College of Mechatronics and Control Engineering, Shenzhen University, 3688 Nanshan Avenue, Shenzhen, 518060 Guangdong, China
[b] George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, 813 Ferst Drive, NW, Atlanta, GA 30332-0405, USA
[c] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Department of Biomedical Engineering, School of Medicine, Shenzhen University, Shenzhen 518060, China

ARTICLE INFO

ABSTRACT

With the development of social media, online social networks offer potential opportunities for firms to analyze user behaviors. Among many, one of the fundamental questions is how to predict product adoption, and the answer to this question lays the foundation for product adoption maximization and demand estimation in large social networks. However, due to the inherent challenges resulting from the dynamic diffusion mechanism in online social networks, such as modeling of activation thresholds and influence probability, differentiating between influence and adoption, and incorporating review content, traditional diffusion models are often not adequate enough to predict product adoption accurately. In order to tackle these challenges, we propose a linear threshold-hurdle model to predict product adoption incorporating social network effects. First, we present a fine-grained activation threshold model based on the five categories of adopters. In addition, we identify three operational factors underlying social network effects, including interaction strength, structural equivalence, and social entity similarity, to model influence probabilities. Furthermore, we distinguish influence spread from adoption spread by introducing a tattle state, in which users express opinions without adopting the product. Finally, we introduce the notion of hurdle to capture the monetary aspect in users' decision making process of product adoption. Based on the proposed linear threshold-hurdle model, two data mining methods based on the rough set technique, namely, decision rules and decomposition trees, are employed to predict product adoption in a large social network. An empirical study of Kindle Fire HD 7 in. tablets is used to illustrate the potential and feasibility of the proposed model. The results demonstrate the predictive power of the proposed model with average F-scores of 89.8% for the week prediction model and of 86.7% for the bi-week prediction model.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

A modern product like an iPhone or iPad works not only because of its inherent industrial and interface design, but also because of the social networks in which it "lives" [14]. With the pervasive connectivity of the Internet and social media, including review sections of online shopping websites (e.g., Amazon.com) and online social networks (e.g., Facebook),

customers become more interconnected and informed when they make product choices. The social network plays a fundamental role as a medium for the spread of information, ideas, and influence among its social entities [34]. For example, the adoption of a new mobile phone among college students, the adoption of a new weed spray in a village, and the adoption of a new drug in a medical society, would die out quickly or spread to a large population due to the information diffusion in the social network [see 50].

In this process, the social entities consider not only the attributes of the products, but also the preferences of other customers in the social networks, or are influenced by the "word of mouth" in the process of new product promotion [43]. These effects can be understood as social network effects and often take place when customers aspire to be like or unlike others, or learn something new about certain products from others. There is little doubt that one can hardly isolate his purchase or usage decisions from his social networks. Such effects often lead to the spread of adoption behavior from one social entity to another in the social network [36].

The increasing availability of social network data has drawn more attention to understand the social network effects on customers' product adoption decisions and adoption maximization [32]. One of the fundamental questions is how to predict product adoption for social entities who have not adopted by now [17]. The answer to this question is not only critical to viral marketing and design with regard to product adoption in social networks [e.g., 2,11], but also vital to applications in demand estimation [e.g., 29], public health [e.g., 13], and politics [e.g., 5], etc. For example, in viral marketing, it is important to identify a set of powerful influencers in the social network as seeds so that the expected number of social entities who adopt the product can be maximized. This is dependent on the reliable prediction of product adoption in the search process of optimal seeds because one needs to predict how likely other entities will adopt if the initially targeted ones adopt [17].

1.1. Technical challenges

(1) **Modeling influence probability:** In order to predict product adoption, it is important to understand how the dynamics of adoption are likely to unfold within the underlying social network. Among many, two types of diffusion models [34], i.e., independent cascade models and linear threshold (LT) models, are proposed. In both types of models, the key problem is how to model the influence probability from one social entity to another. First, in cascade models, the adoption probability is often modeled in an ad hoc way. The adoption process unfolds in discrete steps. Assuming v is an inactive social entity (i.e., non-adopter) at step t, and the probability of v becoming active (i.e., adopters) depends on the influence of v's neighbors who are adopters at step t. It can be calculated as $p_v = 1 - \prod_{u \in N}(1 - p_{v,u})$ [11,12], where N is the set of active neighbors of v, and $p_{v,u}$ is the influence probability of u on v. It can be seen that the adoption probability is calculated based on the assumption that each active neighbor influences v independently. Besides, $p_{v,u}$ is often set as $1/k$ for all $u \in N$, where k is the total number of active neighbors of v [11,12], or a constant, such as 0.1 or 0.01 [34,35]. Such a simple model cannot accommodate the dynamic nature of the adoption process in online social networks, and thus further investigation is needed. Second, like the cascade models, the adoption probability in LT models is also modeled in a simple way. For example, Goyal et al. [22] proposed to learn $p_{v,u}$ based on the past behavior between u and v, specifically, as the ratio of the number of actions propagated from u to v to the total number of actions performed by u. However, no other factors, such as network structures and entity properties, are taken into consideration. For example, it has been proved that socially connected users tend to be similar (i.e., homophily [42]), which can be leveraged to predict product adoption.

(2) **Modeling activation threshold:** In LT models, each social entity v has an activation threshold θ_v, which follows a uniform distribution between 0 and 1. At each discrete step, if the sum of the incoming influence from the active neighbors exceeds the threshold, v will become active. However, each social entity's threshold is randomly selected, which is not consistent with the studies in the domain of innovation diffusion and communication. Rogers [50] identified five categories of adopters in diffusion research, including innovators, early adopters, early majority, late majority, and laggards. Apparently, each category of adopters does not have an activation threshold distributed uniformly between 0 and 1. If each social entity can be grouped into a certain category, its activation threshold can be better specified.

(3) **Distinguishing influence spread from adoption spread:** Another challenge in diffusion models is how to separate influence spread from adoption spread [7]. In LT models, these two concepts are considered as the same. In other words, once a social entity is active, he or she will automatically and unconditionally become an adopter. This is not necessarily true in reality. In both sociology and marketing literature [33,50], it has been pointed out that influence and adoption are two different concepts. In the influence stage, a social entity becomes aware of the product and gets familiar with its features, which is often used as a proxy for adoption [7]. The influence spread is often taking place in an epidemic-like manner, which is articulated by the typical diffusion models. However, the actual adoption also depends on other factors, such as price and individual's valuation of the product, which are not captured in classic diffusion models [41].

(4) **Incorporating both positive and negative reviews:** When a social entity adopts a product, it is assumed that he or she will influence others to adopt it as well in a positive way in the classic diffusion models [7]. Obviously, this is not entirely true. A certain percentage of users will give negative reviews about the product, if it cannot satisfy their user needs. Therefore, it is important to incorporate negative product reviews that discourage other users from adopting the product in the diffusion process. Another assumption may not hold is that only adopters can share their user