

## 基于线性阈值模型的影响力传播权重学习

郭静<sup>\*①②</sup> 曹亚男<sup>②</sup> 周川<sup>②</sup> 张鹏<sup>②</sup> 郭莉<sup>②</sup>

<sup>①</sup>(北京邮电大学计算机学院 北京 100876)

<sup>②</sup>(中国科学院信息工程研究所 北京 100093)

**摘要:** 度量用户间影响力对商品的营销和推广具有重要作用。然而,已有工作通常假设用户之间的相互影响行为是独立的,忽略了影响力在传播过程中具有的累积效应。为解决此问题,该文在线性阈值模型的框架下,提出一种影响力传播权重的计算方法。该方法将社交网络中用户的历史行为日志看作样本,借鉴最大似然估计的思想对用户间影响力学习问题建模,并设计一种优化的粒子群算法对问题求解。实验使用真实数据验证了该方法的有效性。

**关键词:** 社交网络; 线性阈值模型; 影响力分析; 传播权重

中图分类号: TP393; TP399

文献标识码: A

文章编号: 1009-5896(2014)08-1804-06

DOI: 10.3724/SP.J.1146.2014.00090

## Influence Weights Learning under Linear Threshold Model in Social Networks

Guo Jing<sup>①②</sup> Cao Ya-nan<sup>②</sup> Zhou Chuan<sup>②</sup> Zhang Peng<sup>②</sup> Guo Li<sup>②</sup>

<sup>①</sup>(School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China)

<sup>②</sup>(Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China)

**Abstract:** Quantizing the influence propagation weights between users plays an important role in commodity marketing and promoting in social networks. However, most of current studies assume the mutual behaviors between users to influence each other are independent, while overlooked the accumulative effect in influence propagation process. To fill this gap, this study proposes an influence weights learning approach under the framework of the linear threshold model. With a log of past propagations of involved users in social networks, the study formulize an objective function on the basis of maximum likelihood estimation for the proposed problem, and presents a particle swarm optimization algorithm according to the objective function. Experimental results on real-world datasets validate the effectiveness of the proposed approach.

**Key words:** Social networks; Linear Threshold (LT) model; Influence analysis; Propagation weights

### 1 引言

随着互联网技术的不断发展,社交网络成为人们交流互动的新场所<sup>[1]</sup>。在社交网络中,用户通过影响力的传播作用影响着网络中其他用户的情感、观点和行为。因此,用户间的影响力度量问题引起大量研究<sup>[2-12]</sup>。早期的研究往往基于社交网络的拓扑信息,利用常数赋值法、度平均计算法<sup>[2-5]</sup>、随机赋值法<sup>[2,6]</sup>来估算用户间的影响力。然而,这些方法缺少系统的学习与分析,且忽略了用户历史行为的作用。为此,一些研究工作以用户的历史行为为基础,利用统计和模型学习的方法来度量用户之间的

影响力。例如:文献[7]在用户历史行为记录的基础上,利用期望最大化(Expectation-Maximization, EM)算法估算用户间的影响力。文献[8]以 Jaccard 系数和伯努利分布为基础,提出若干概率模型来建模用户间的影响力学习问题。文献[9]从主题模型的角度研究传播影响力,并计算出基于主题有效的独立级联模型(Topic-aware Independent Cascade, TIC)的相关参数。然而,此类方法在建模时,都假设用户的激活行为是相互独立的。在现实中,用户的朋友可能共同影响其做出某一行为,所以上述方法不能全面度量用户之间的影响力。

为解决上述问题,本文以线性阈值(Linear Threshold, LT)模型<sup>[4,13]</sup>为基础来学习用户之间的影响力。为此,我们将会面临下列挑战。首先,LT模型以权值来度量用户间影响力,缺少相应的概率描述,所以无法直接使用最大似然估计方法对问题建

2014-01-15 收到, 2014-04-24 改回

中国科学院战略性先导科技专项(XDA06030200), 中国科学院信息工程研究所科研项目(Y3Z0062101)和国家自然科学基金青年基金(61003167)资助课题

\*通信作者: 郭静 guojing@nelmail.iie.ac.cn

模。其次，由于用户间影响力度量与多个邻居用户的历史行为相关，所以每个用户都需要建立不同目标函数，如何对不同的目标函数有效求解是一个难点。

为此，本文基于 LT 模型，利用最大熵原理估计激活阈值的概率密度函数；并以此计算邻居激活给定用户的概率，并将用户的历史行为日志看作样本，借鉴最大似然估计的思想对用户间影响力学习问题建模；最后，设计一种改进的粒子群算法来求解问题。该算法根据问题目标函数和约束的特点，通过问题映射、适应度函数建立、越界阻止、动态参数设置和最优粒子变异等优化策略，有效地学习用户间影响力。

## 2 相关预备知识

LT 模型是一种能够刻画“影响力累计特性”的传播模型。考虑到现实中用户可能被其若干朋友共同影响，本文基于 LT 模型来学习用户间的影响力。本节先回顾相关预备知识，表 1 是相关数学符号表示。

表 1 对应数学符号表示

$G = (V, E)$	社交网络的结构图， $V, E$ 分别为点集和边集
$w$	网络中节点 $w$
$N(u)$	特定用户 $u$ 的邻居节点集合
$\theta_u$	LT 模型下，节点 $u$ 的激活阈值
$b_{u,w}$	LT 模型下，沿有向边 $\vec{u,w}$ 的激活权重

在 LT 模型中，每个节点存在激活或未激活状态，且只处于两者之一。每条边  $(w, u) \in E$  都被赋予  $b_{w,u}$ ，以表示  $w$  对  $u$  的影响力， $\theta_u$  反映  $u$  被邻居激活的可能性。当激活邻居的所有权重和超过该点阈，如式(1)，则该节点被激活；否则激活失败。

$$\sum_{w \in N(u)} b_{w,u} \geq \theta_u \quad (1)$$

另外，对每个节点  $u$ ，所有邻居对其产生的影响力权重之和不能超过 1。

$$\sum_{w \in N(u)} b_{w,u} \leq 1 \quad (2)$$

## 3 基于 LT 模型的影响力传播权重学习方法

### 3.1 基于最大熵的激活概率计算

LT 模型以权值来度量用户之间的影响力，缺少相应的概率描述，无法直接借鉴最大似然估计的思想来对问题进行建模。为此，本文首先根据 LT 模型的特点来估计邻居激活用户的概率。

在 LT 模型中，邻居是否能激活给定用户，是由邻居对给定用户的影响力和激活阈值所共同确定

的。因此在影响力确定的条件下，邻居激活用户的概率是由激活阈值  $\theta_u$  所决定的。用户是否发生行为受多种因素影响，所以激活阈值具有很强的不确定性。为刻画这种不确定性，本文以最大熵原理来估计激活阈值的概率密度，并以此为基础来计算邻居激活用户的概率。

**定理 1** 已知激活阈值  $\theta_u$  的取值范围在  $[0,1]$  之间，且  $\theta_u$  的概率密度函数  $f(\theta_u)$  满足等式(4)，则激活阈值  $\theta_u$  的最大熵分布的概率密度函数满足  $f(\theta_u) = 1$ 。

**证明** 根据最大熵原理<sup>[14]</sup>，在满足约束的条件下，熵值最大的分布就是符合实际的分布。由于没有先验知识，仅知道  $\theta_u$  的取值范围在  $[0,1]$  之间，所以求解  $\theta_u$  的概率密度  $f(\theta_u)$ ，就是求解下列约束优化问题。其中， $H(\theta_u)$  是激活阈值  $\theta_u$  的熵值， $C$  为常数。

$$\text{Max } H(\theta_u) = -C \int_0^1 f(\theta_u) \ln f(\theta_u) d\theta_u \quad (3)$$

$$\int_0^1 f(\theta_u) d\theta_u = 1 \quad (4)$$

为此，首先用拉格朗日乘子法将约束加入目标形成新目标函数式(5)，然后使用求导方法计算问题的最优值。

$$\text{Max } H(\theta_u) = -C \int_0^1 f(\theta_u) \ln f(\theta_u) d\theta_u + \lambda \left( \int_0^1 f(\theta_u) d\theta_u - 1 \right) \quad (5)$$

通过目标函数式(5)对  $f(\theta_u)$  求导，可得式(6)。

$$-C \int_0^1 (\ln f(\theta_u) + 1) d\theta_u + \lambda \int_0^1 d\theta_u = 0 \quad (6)$$

$$f(\theta_u) = e \left( \frac{\lambda}{C} - 1 \right) \quad (7)$$

将式(7)代入式(4)可得  $f(\theta_u) = 1$ 。证毕

在定理 1 的基础上，本文利用激活阈值  $\theta_u$  的概率密度函数  $f(\theta_u)$  来计算邻居激活用户的概率。假设邻居对用户  $u$  的影响力为  $b_{w,u}$ ，则邻居成功激活用户的概率  $P \left( \sum_{w \in N(u)} b_{w,u} \geq \theta_u \right)$  可由式(8)计算获得，邻居激活用户失败的概率  $P \left( \sum_{w \in N(u)} b_{w,u} < \theta_u \right)$  可由式(9)计算获得。

$$P \left( \sum_{w \in N(u)} b_{w,u} \geq \theta_u \right) = \int_0^{\sum_{w \in N(u)} b_{w,u}} f(\theta_u) d\theta_u = \sum_{w \in N(u)} b_{w,u} \quad (8)$$

$$P \left( \sum_{w \in N(u)} b_{w,u} < \theta_u \right) = 1 - \int_0^{\sum_{w \in N(u)} b_{w,u}} f(\theta_u) d\theta_u = 1 - \sum_{w \in N(u)} b_{w,u} \quad (9)$$

### 3.2 问题建模

根据用户的历史行为日志，本文借鉴最大似然估计的思想，对用户影响力学习问题建模。即：将日志信息看作是基于 LT 模型产生的样本，利用用户影响力取值使样本观测的概率最大的思想对问题建模。本文使用三元组{用户-事件-时间}来表现日志中用户在事件上的激活状态， $U$  表示用户集合， $D$  表示事件集合， $T$  表示时间集合。针对历史日志中的某事件  $D_i (i = \{1, 2, \dots, |D|\}) \in D$ ，用户  $u$  的状态分两种情况讨论：

(1)  $u$  激活：若其所有邻居的激活时间都晚于  $u$ ，则表示  $u$  自发参与到  $D_i$  的传播中；若存在邻居的激活时间都早于  $u$ ，则表示  $u$  受到邻居影响才参与到  $D_i$  的传播中，它是邻居成功激活用户的样本。

(2)  $u$  未激活：若其所有邻居都处于未激活状态，则表示  $u$  既没有受到来自邻居的影响，也没有自发地参与；若存在激活的邻居，则表示邻居对  $u$  的激活失败，它是邻居激活用户失败的样本。

根据日志里的用户信息，本文利用式(10)描述用户间影响力学习的目标函数。即

$$\left. \begin{aligned} L(\mathbf{X}, \mathbf{D}) = \text{Max} \quad & \prod_{D_i \in \mathbf{D}, i = \{1, 2, \dots, |D|\}} Q(u, D_i) \\ \text{s.t.} \quad & \sum_{w \in \mathbf{N}(u)} b_{w,u} \leq 1 \\ & b_{w,u} \geq 0 \end{aligned} \right\} \quad (10)$$

其中， $L(\mathbf{X}, \mathbf{D})$  是本问题的目标， $Q(u, D_i)$  是用户  $u$  在事件样本  $D_i$  上的状态，其取值可由式(11)来表示。

$$Q(u, D_i) = \begin{cases} P(u, D_i), & u \text{ 处于激活状态} \\ F(u, D_i), & u \text{ 处于未激活状态} \end{cases} \quad (11)$$

$P(u, D_i)$  和  $F(u, D_i)$  分别对应  $u$  在  $D_i$  中处于激活和未激活的两种状态，如式(12)和式(13)所示。

$$P(u, D_i) = \begin{cases} 1, & t(\forall v \in \mathbf{N}(u, D_i)) \geq t(u) \\ \sum_{w \in \mathbf{N}(u, D_i)} b_{w,u}, & t(\exists v \in \mathbf{N}(u, D_i)) < t(u) \end{cases} \quad (12)$$

$$F(u, D_i) = \begin{cases} 1, & \mathbf{N}(u, D_i) = \Phi \\ 1 - \sum_{w \in \mathbf{N}(u, D_i)} b_{w,u}, & \mathbf{N}(u, D_i) \neq \Phi \end{cases} \quad (13)$$

其中， $\mathbf{N}(u, D_i)$  表示在  $D_i$  中处于激活状态的邻居集合， $\mathbf{N}(u, D_i) \subseteq \mathbf{N}(u)$ ； $t(\cdot)$  表示用户被激活的时间， $\Phi$  表示空集。对每个  $u$  而言，所有邻居对  $u$  产生的影响力和不能超过 1，所以目标函数应满足约束条件。

### 3.3 求解算法

本文根据问题目标和约束条件，设计一种改进粒子群算法 (Improved Particle Swarm Optimization Algorithm for Influence Weights Learning, IWL-IPSO) 来求解问题。其优化策略包

括：(1)问题映射，(2)适应度函数建立，(3)越界阻止，(4)动态参数设置，(5)最优粒子变异等。

**3.3.1 粒子群算法回顾** 粒子群算法<sup>[15]</sup>是一种基于群体智能的启发式全局搜索算法。它从随机解出发，迭代搜索问题的最优解。即根据粒子的最好位置  $p_{\text{best}}(t)$  和粒子群中最优粒子的位置  $p_{\text{gbest}}(t)$ ，使用式(14)和式(15)来更新位置信息。

$$v_{\text{id}}^{t+1} = wv_{\text{id}}^t + c_1r_1(p_{\text{best}}(t) - x_{\text{id}}^t) + c_2r_2(p_{\text{gbest}}(t) - x_{\text{id}}^t) \quad (14)$$

$$x_{\text{id}}^{t+1} = x_{\text{id}}^t + v_{\text{id}}^{t+1} \quad (15)$$

其中，粒子位置代表问题的解； $t$  是当前迭代次数； $x_{\text{id}}^t$  是第  $i$  个粒子经第  $t$  次迭代后，在第  $d$  维空间的位置； $v_{\text{id}}^t$  是当前速度； $c_1$  和  $c_2$  是学习因子，代表粒子向局部和全局最优位置移动的趋势； $r_1$  和  $r_2$  是随机数； $w$  是粒子的惯性权重。根据本问题目标和约束条件，本文采用下列优化步骤来改进传统粒子群算法的搜索步骤。

**3.3.2 问题的粒子映射** 问题的合理映射是粒子群算法求解问题的前提，由于线性阈值模型中用户的状态受邻居的影响，所以本文将邻居用户  $w \in \mathbf{N}(u)$  对用户  $u$  的影响力  $b_{w,u}$  的组合表示成一个粒子，具体如图 1 所示。

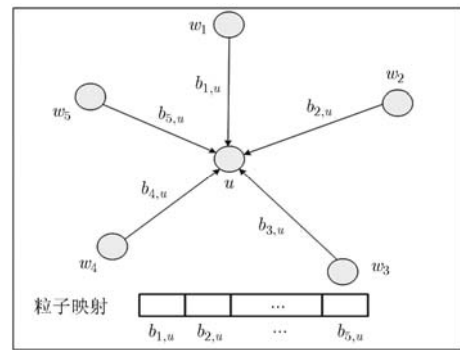


图1 问题的粒子映射

在图1中，每个粒子由  $n(u)$  个位置信息  $(x_1, x_2, \dots, x_{n(u)})$  构成，其中， $n(u)$  是用户  $u$  的邻居集合  $\mathbf{N}(u)$  中的节点数目；粒子的每个位置  $x_i, (i \in \{1, 2, \dots, n(u)\})$  表示邻居用户  $w_i \in \mathbf{N}(u)$  对用户  $u$  的影响力  $b_{w_i,u}$ 。

**3.3.3 适应度函数设计** 适应度函数是粒子群算法学习用户间影响力的关键。由于  $L(\mathbf{X}, \mathbf{D})$  的计算结果与历史事件的数量  $|D|$  相关，且  $b_{w,u}$  满足约束条件  $0 \leq b_{w,u} \leq 1$  和  $\sum_{w \in \mathbf{N}(u)} b_{w,u} \leq 1$ ，故当  $|D|$  较大时，有  $\lim_{|D| \rightarrow \infty} L(\mathbf{X}, \mathbf{D}) = 0$ ，难以评价方案适应度；此外，不

满足约束方案中也可能包含着最优解的影响力取值。为此，本文设计一种新的适应度函数，即采用开 $|D|$ 次方的形式来增大适应度的取值，并使用分段处理策略来综合评价影响力权重，如式(16)。

$$\text{Fit}(\mathbf{X}) = \begin{cases} 0.5 + 0.5 \times \sqrt{|D|} \sqrt{L(\mathbf{X}, \mathbf{D})}, & \sum_{w \in N(u)} b_{w,u} \leq 1 \\ 0.5 \times \sqrt{|D|} \sqrt{L(\mathbf{X}, \mathbf{D})} - \sum_{w \in N(u)} b_{w,u} + 1, & \text{其它} \end{cases} \quad (16)$$

其中， $L(\mathbf{X}, \mathbf{D})$ 是问题的目标函数， $b_{w,u}$ 是粒子搜索到的影响力传播权重。

**3.3.4 约束违背控制** 由于LT模型中的用户影响力取值范围在 $[0,1]$ ，传统粒子群算法没有直接方法来限制粒子位置，使得粒子在搜索过程容易飞离搜索区域，产生无效的影响力权重。为了使更多粒子在有效的区域中搜索，本文在位置更新式(15)中加入动力参数设置，具体如式(17)所示。

$$x_{id}^{t+1} = \lambda x_{id}^t + (1 - \lambda) v_{id}^{t+1} \quad (17)$$

其中， $\lambda$ 是动力学参数，其取值满足 $0 < \lambda < 1$ 。

**3.3.5 动态参数设置** 在用户影响力学习的过程中，算法的全局和局部搜索能力共同决定了学习的效果。文献[16]指出，当惯性权重取值 $w$ 较小时，算法将拥有较强的局部搜索能力，有利于粒子收敛。当惯性权重取值 $w$ 较大，将增强粒子的探索能力，使得算法的全局搜索能力得到提高。为平衡二者的关系，本文以递减函数来改变 $w$ 的取值，让 $w$ 的取值随着迭代过程逐渐递减，其函数表达如式(18)所示。

$$w = w_{\min} + (w_{\max} - w_{\min})(t/T) \quad (18)$$

其中， $w_{\max}$ 是惯性权重系数取得的最大值， $w_{\min}$ 是惯性权重系数取得的最小值， $T$ 是算法的最大迭代次数， $t$ 是当前的迭代次数。在式(18)的作用下，粒子在算法的执行初期，将更加注重信息的获取；而在算法的执行后期，它将更加倾向于对当前最优方案方向进行搜索，从而提高算法的收敛性。

**3.3.6 最优粒子变异** 由于全局最优粒子的位置满足 $p_{\text{gbest}}(t) = p_{\text{best}}(t) = x_{id}^t$ ，其位置更新仅由惯性速度 $wv_{id}^t$ 决定。即：该粒子将一直按照惯性方向运动，运动具有盲目性。特别当 $v_{id}^t = 0$ 时，该粒子将保持当前位置不变，可能导致算法收敛于局部最优解，出现早熟现象。为克服此问题，本文对最优粒子进行变异操作。其思想就是随机选择粒子的任意维位置，利用式(19)对粒子在该维空间中的位置进行随机搜索，以获得更好的取值。

$$x_{\text{gbest},d}^{t+1} = p_{\text{gbest},d}(t) + \delta \quad (19)$$

其中， $\delta$ 表示全局最优粒子在位置 $p_{\text{gbest},d}(t)$ 进行随机搜索的范围，其取值与粒子其他位置上的取值相关，

其取值满足式(20)。

$$\delta = \begin{cases} r_d \in [-p_{\text{gbest},d}(t), 0], & \sum_{d=1}^n p_{\text{gbest},d}(t) > 1 \\ c_d \in \left[-p_{\text{gbest},d}(t), 1 - \sum_{d=1}^n p_{\text{gbest},d}(t)\right], & \text{其它} \end{cases} \quad (20)$$

由式(20)可以看出，当 $\sum_{d=1}^n p_{\text{gbest},d}(t) > 1$ 时，全局最优粒子向减小约束违背的方向进行搜索；而当粒子满足约束条件时，它将根据粒子在其他位置上的取值情况，在满足约束的条件下进行随机搜索。

## 4 实验

为验证本方法的合理性，在多次实验的结果上，以方法执行时间、用户间影响力的适应度为评价指标来综合分析IWL-IPSO算法的性能。其中，数据集来自文献[17]，对比算法的描述情况：

(1)IWL-PSO(Particle Swarm Optimization algorithm for Influence Weights Learning): 该算法直接利用传统粒子群算法来学习用户间的影响力。

(2)IWL-Degree(Degree algorithm for Influence Weights Learning): 该算法直接根据社交网络中邻居的数目计算用户间的影响力。

(3)IWL-PSOA(Particle Swarm Optimization algorithm A for Influence Weights Learning): 该算法在IWL-IPSO的基础上，通过去除约束违背控制操作来学习用户间的影响力。

(4)IWL-PSOB(Particle Swarm Optimization algorithm B for Influence Weights Learning): 该算法在IWL-IPSO的基础上，通过去除动态参数设置操作来学习用户间的影响力。

(5)IWL-PSOC(Particle Swarm Optimization algorithm C for Influence Weights Learning): 该算法在IWL-IPSO的基础上，通过去除最优粒子变异策略来学习用户间的影响力。

### 4.1 方法有效性验证

为验证方法的有效性，本次实验将从数据集中选择拥有不同历史行为记录的用户为分析对象，利用IWL-IPSO, IWL-PSO和IWL-Degree来学习邻居用户对他们的影响力，并根据算法的执行时间和用户间影响力的适应度来分析方法在不同历史行为记录下的性能。其中，被选目标用户的详细信息如表2所示。

图2和图3是IWL-IPSO, IWL-PSO和IWL-Degree在20次实验中对用户间影响力学习所使用的平均时间及获得的平均适应度。从图2中可以看出，在邻居数目相同的情况下，IWL-IPSO的执行时间

表2 被选用户的情况

用户ID	309798	325985	11187	73588	179577	79988	159549
用户的邻居个数	60	60	60	60	60	60	60
历史信息数量	55	214	422	653	806	1104	1407

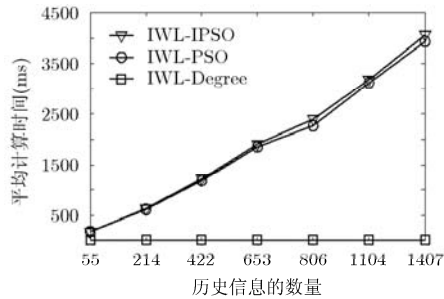


图2 不同算法执行的平均时间

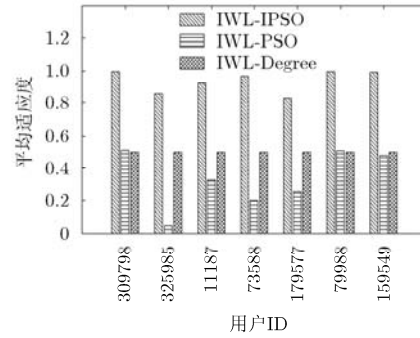


图3 不同算法学习用户间影响力获得的平均适应度

随历史信息数量的增长呈线性增长。从图3可以看出，IWL-IPSO在不同历史行为记录下均取得较好的适应度。

### 4.2 方法的性能分析

为进一步分析方法的性能，实验选择具有不同邻居数目、不同历史行为记录的用户为分析对象，通过去除IWL-IPSO的优化步骤来分析它们对学习效果的影响。被选择的用户信息如表3所示，实验结果如表4和表5所示。

从表3和表4可以看出，每个改进步骤都可以提高影响力学习的适应度，且步骤的花费时间较方法整体时间而言相对较少。其中，IWL-PSOA在使用约束违背控制策略后，其学习影响力的适应度较原

表3 性能分析所选择用户的情况

用户ID	135203	146418	162911	317607	320209
用户的邻居个数	5	10	18	24	31
历史信息数量	32	100	51	165	2286

表4 不同算法学习用户间影响力获得的平均适应度

用户ID	IWL-IPSO	IWL-PSOA	IWL-PSOB	IWL-PSOC
135203	0.872383	0.872224	0.872235	0.863409
146418	0.878782	0.877155	0.877464	0.842794
162911	0.660425	0.655434	0.655712	0.641625
317607	0.939727	0.933116	0.938550	0.830155
320209	0.994722	0.994425	0.992146	0.987525

表5 不同方法在影响力学习的计算时间(ms)

用户ID	IWL-IPSO	IWL-PSOA	IWL-PSOB	IWL-PSOC
135203	53	53	51	50
146418	167	167	166	159
162911	103	101	100	97
317607	327	326	322	316
320209	5017	4951	4938	4803

方法提高了0.34%，这是因为该策略对传统的更新值进行了缩放，尽可能让用户间的影响力满足约束。IWL-PSOB在使用动态参数设置策略后，其学习影响力的适应度较原方法提高了0.25%，这是因为该策略平衡了算法的局部和全局搜索的能力，粒子在算法的执行初期，更加注重信息的获取；而在算法的执行后期，更加倾向于对当前最优方案方向进行搜索，从而提高算法的收敛性。IWL-PSOC在使用最优粒子变异策略后，其学习影响力的适应度较原方法提高了4.43%。这是因为在传统的粒子群算法中，最优粒子的运动具有盲目性，可导致算法收敛于局部最优解，出现早熟现象，而最优粒子变异策略可以有效克服上述问题。

## 5 结束语

本文在线性阈值模型的框架下，提出一种影响力传播权重的计算方法。该方法基于社交网络中用户的历史行为日志信息，利用优化的粒子群算法来学习用户间的影响力。同时，本文基于真实的社交网络数据和相关用户的历史行为日志，实验证明了本文方法的有效性。

## 参 考 文 献

- [1] Park B, Lee K, and Kang N. The impact of influential leaders in the formation and development of social networks[C]. Proceedings of the 6th International Conference on Communities and Technologies(C&T), New York, USA, 2013: 8-15.
- [2] Chen Wei, Wang Chi, and Wang Ya-jun. Scalable influence maximization for prevalent viral marketing in large-scale social networks [C]. Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Washington, USA, 2010: 1029-1038.
- [3] Chen Wei, Wang Ya-jun, and Yang Si-yu. Efficient influence maximization in social networks[C]. Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Paris, France, 2009: 199-208.
- [4] Kempe David, Kleinberg Jon, and Tardos Eva. Maximizing the spread of influence through a social network[C]. Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Washington, USA, 2003: 137-146.
- [5] Zhou Chuan, Zhang Peng, Guo Jing, *et al.* UBLF: an upper bound based approach to discover influential nodes in social networks[C]. Proceedings of the IEEE International Conference on Data Mining (ICDM), Dallas, USA, 2013: 907-916.
- [6] Jung Kyomin, Heo Wooram, and Chen Wei. IRIE: scalable and robust influence maximization in social networks[C]. Proceedings of the IEEE International Conference on Data Mining (ICDM), Brussels, Belgium, 2012: 918-923.
- [7] Saito Kazumi, Nakano Ryohei, and Kimura Masahiro. Prediction of information diffusion probabilities for independent cascade model[C]. Proceedings of the International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES), Zagreb, Croatia, 2008: 67-75.
- [8] Goyal Amit, Bonchi Francesco, and Lakshmanan Laks V S. Learning influence probabilities in social networks[C]. Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM), New York, USA, 2010: 241-250.
- [9] Barbieri Nicola, Bonchi Francesco, and Manco Giuseppe. Topic-aware social influence propagation models[C]. Proceedings of the IEEE International Conference on Data Mining (ICDM), Brussels, Belgium, 2012: 81-90.
- [10] Liu Lu, Tang Jie, Han Jia-wei, *et al.* Learning influence from heterogeneous social networks[J]. *Data Mining and Knowledge Discovery*, 2012, 25(3): 511-544.
- [11] Konstantin Kutkov, Albert Bifet, Francesco Bonchi, *et al.* STRIP: stream learning of influence probabilities[C]. Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Chicago, USA, 2013: 275-283.
- [12] Goyal Amit, Bonchi Francesco, and Lakshmanan Laks V S. A data-based approach to social influence maximization[C]. Proceedings of the International Conference on Very Large Data Bases (VLDB), Istanbul, Turkey, 2012: 73-84.
- [13] Goyal Amit, Lu Wei, and Lakshmanan Laks V S. Simpath: an efficient algorithm for influence maximization under the linear threshold model[C]. Proceedings of the IEEE International Conference on Data Mining (ICDM), Vancouver, Canada, 2011: 211-220.
- [14] Jaynes E T. Information theory and statistical mechanics[J]. *Physical Review*, 1957, 106(4): 620-630.
- [15] Eberhart Russell and James Kennedy. A new optimizer using particle swarm theory[C]. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, 1995: 39-43.
- [16] Shi Yuhui and Eberhart Russell. A modified particle swarm optimizer[C]. Proceedings of IEEE World Congress on Computational Intelligence, Anchorage, USA, 1998: 69-73.
- [17] Liu Lu, Tang Jie, Han Jia-wei, *et al.* Mining topic-level influence in heterogeneous networks[C]. Proceedings of the Conference on Information and Knowledge Management (CIKM), Toronto, Canada, 2010: 199-208.
- 郭 静： 女，1982年生，博士生，研究方向为社会计算、数据挖掘。
- 曹亚男： 女，1985年生，助理研究员，研究方向为Web语义抽取、文本挖掘。
- 周 川： 男，1984年生，博士后，研究方向为概率统计。
- 张 鹏： 男，1982年生，副研究员，研究方向包括社会计算、数据挖掘。
- 郭 莉： 女，1969年生，教授，研究方向为网络信息安全。