

理性背叛的推荐合作激励机制

金星, 李明楚, 孙晓梅, 郭成

(大连理工大学 软件学院, 辽宁 大连 116620)

E-mail: jinxing@mail.dlut.edu.cn

摘要: 为了促进协作系统中用户的合作行为, 激励机制得到了广泛的使用. 然而, 现有的激励机制往往存在无条件合作策略占优互惠策略的现象, 进而抑制了合作的涌现. 为了解决这一问题, 本文在推荐激励模型上进一步考虑了用户的理性背叛行为. 以演化博弈为框架, 研究了理性背叛机制在全局平均学习和当前最优学习两种模式下的策略演化特性. 结合实际场景, 本文还研究了在非完美推荐下理性背叛机制的鲁棒性问题, 并且基于余弦相似度提出了一种策略识别方案. 最后, 通过大量的数值实验与仿真实验, 验证了理性背叛机制的理论特性, 也展示了该机制在促进合作方面的有效性能.

关键词: 合作; 激励机制; 理性背叛; 演化博弈; 协作系统

中图分类号: TP393

文献标识码: A

文章编号: 1000-1220(2020)02-0244-08

Recommendation Incentive Mechanisms with Rational Defection for Promoting Cooperation

JIN Xing, LI Ming-chu, SUN Xiao-mei, GUO Cheng

(School of Software, Dalian University of Technology, Dalian 116620, China)

Abstract: To promote cooperation in collaborative systems, incentive mechanisms have been widely used. However, in existing incentive mechanisms, it is found that the unconditional cooperation strategy always dominates the reciprocity strategy, which inhibiting the emergence of cooperation. To address this issue, based on the recommendation incentive model, we further consider the defection behaviors of rational users'. And then, by using the evolutionary game theory, we study the population dynamics of the proposed rational defection model under two different learning mechanisms: GMLM (Global Mean Learning Mechanism) and CBLM (Current Best Learning Mechanism). In order to incorporate practical application, firstly, we study the robustness of our model under the scenario of imperfect recommendation, and then, based on the cosine similarity we propose a method to distinguish between strategies. At last, extensive numerical and simulation experiments are conducted to validate the theoretical properties of our model and illustrate the sufficiency of our model in promoting cooperation.

Key words: cooperation; incentive mechanisms; rational defection; evolutionary game; collaborative systems

1 引言

近年来,随着互联网和移动网络的飞速发展,基于网络合作的协作系统已大量出现在了我们的生活中. 在这类系统中,虽然每一个个体的能力都是有限的,但是通过个体之间的合作行为,人们可以实现一系列复杂困难的任务. 然而,由于个体理性的存在,参与者往往会选择背叛行为,从而造成了合作困境. 一个典型的例子就是 P2P 文件共享系统里用户的搭便车现象^[1].

为了促进用户的合作行为,保证协作系统的有效运行,引入激励机制已经成为了必不可少的选择. 通过研究之前的文献,可以发现:在考虑策略花费这一实际场景下,互惠策略会占优无条件背叛策略,无条件背叛策略会占优无条件合作策略,无条件合作策略会占优互惠策略^[2-4],这种互相占优关系会导致无条件合作策略抑制合作产生的这一奇怪现象. 为了解决上述现象,本文在推荐合作激励模型^[4]基础上引入了用

户的理性背叛机制,最后理论和实验都证明了该机制在促进合作方面的有效性能.

2 相关工作

为了解决合作困境,研究者们进行了大量的激励机制研究. 根据之前的文献,激励机制研究主要包含两方面的内容:金钱奖励机制和互惠机制. 在金钱奖励机制中,合作者可以通过提供服务获取物质报酬,因其高效性和简单性,该机制已广泛应用于实际的协作系统中,例如:滴滴网约车平台,亚马逊在线工人平台,等等. 在很多没有设置金钱奖励的协作系统里,互惠机制发挥了极其重要的作用. 根据交互对象的特点,互惠机制可以分为:直接互惠和间接互惠. 在直接互惠中,交互双方是保持不变的,其核心思想是“我帮助你,你帮助我”. 一报还一报策略 (Tit-for-Tat strategy, TFT)^[5] 是目前最著名的直接互惠策略, TFT 策略使用者首先会采用合作行为,而在

之后的交互中会采取对手上一轮的行为. 为了增加 TFT 策略的鲁棒性, TF2T (Tit-for-Two-Tats)^[6] 策略和 GTFT (Generous TFT)^[7] 策略随后也被提了出来. 在 Nowak 的文章^[8]中, 赢留输改 (Win-Stay Lose-Shift, WSLS) 策略被提了出来, 该策略使用者会保留一个收益期望值, 如果当前收益不低于该期望值, 则保留当前行为, 否则, 就更改当前行为, 实验证明 WSLS 取得了比 TFT 更好的性能. 在间接互惠中, 交互双方是动态变化的, 其核心思想是“我帮助别人, 别人帮助我”. 针对大规模 P2P 文件共享场景, Feldman 提出了 Mirror 策略^[9], 当该策略使用者接收到一个服务请求后, 他/她会计算该交互对手为别人提供服务的概率, 然后以相同的概率为该交互对手提供服务. 随后, Zhao 等人^[2]提出了 Proportion 策略, 该策略使用者会计算交互对手提供的服务数量与其获得的服务数量的比例, 并以该比例值作为自己与他人进行合作的概率, 理论和实验证明, 相比 Mirror 策略, Proportion 策略能够更好地促进合作. Wang 等人^[3]研究了花费对 Proportion 策略性能的影响. Li 等人^[4]提出了基于推荐的合作激励策略 (Recommendation Incentive Mechanism, RIM), 该策略使用者在推荐机制的帮助下提高了自身与合作性节点交互的可能性, 进而可以获得更高的收益, 理论和实验证明, 相比 WSLS, Mirror 和 Proportion 策略, RIM 策略具有更好地促进合作的性能. 在文献[4]的基础上, 本文进一步引入了理性背叛机制, 以解决互惠策略被无条件合作策略所抑制的这一问题, 进而希望可以进一步促进合作的涌现.

另一方面, 网络结构对参与者合作行为的影响目前也引起了广泛的关注, 研究者们发现在网格网络^[10], 无标度网络^[11], 小世界网络^[12]等结构化的网络中合作行为可以得到一定程度的促进.

3 模型构建

3.1 演化博弈

Maynard Smith 在 1982 年提出了演化博弈理论模型^[13], 相比于传统的博弈论, 它具有以下两个特性:

1) 考虑了参与者的有限理性. 由于在现实的复杂场景里, 信息往往不能被完全感知, 因此, 参与者无法事先就做出最优的、最理性的策略决策, 这就是参与者的有限理性假设.

2) 动态的博弈结果展示. 在传统的博弈论中, 博弈结果往往是通过纳什均衡进行体现, 然而这是一个静态的结果, 有时人们对博弈结果的生成过程也很感兴趣. 在演化博弈中, 人们不仅可以获得博弈的最终结果, 也可以了解结果的产生过程, 从而对博弈模型有更多的了解.

基于以上两个特点, 本文将使用演化博弈作为模型的基础框架来描述大规模协作系统里用户的交互过程.

3.2 推荐机制框架

基于用户交互的反馈信息, 笔者之前文献[4]通过 Eigen-trust 算法构建了一个合作性用户推荐系统, 使用该系统的用户能够以一个更高的概率与高合作度用户进行交互, 进而提高了获得服务的可能性.

3.3 策略集

假设系统中存在 N 个用户, 每一个用户都会从策略集 S

中选择一个策略 s_i 作为自己与他人进行交互时的行为准则. 在演化博弈中, 使用了相同策略的参与者可以统称为一个种群 (Population).

具体而言, 本文模型考虑了下列三个纯策略.

无条件合作策略 (ALLways Cooperate, ALLC): 该策略使用者会一直为交互对手提供服务, 模型中将该策略记为 s_1 .

无条件背叛策略 (ALLways Defect, ALLD): 该策略使用者不会为其他用户提供服务, 模型中将该策略记为 s_2 .

推荐机制策略 (Recommendation, R): 该策略是一种有条件的合作策略, 他/她会根据交互对象的类型决定自己的行为. 当与 ALLD 策略用户进行交互时, 他/她会选择拒绝为其提供服务. 当与其他 R 策略用户进行交互时, 他/她会选择提供服务. 当与 ALLC 策略用户进行交互时, 他/她会以 f ($f \in [0, 1]$) 的概率为其提供服务. 当 $f = 1$ 时, 就是笔者之前模型^[4]里考虑的场景, 此时会出现无条件合作策略抑制合作涌现的现象. 本文考虑了推荐合作机制中的理性背叛场景, 即 $f \in [0, 1]$. 特别地, 当 $f = 0$ 时, R 策略将不会给 ALLC 策略提供服务. 模型中将该策略记为 s_3 .

基于上述策略设定, 本文引入一个 3×3 矩阵 $P = [p_{ij}]$ 来描述策略间的交互行为:

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ f & 0 & 1 \end{bmatrix}$$

其中元素 p_{ij} 表示策略 s_i 为策略 s_j 提供服务的概率.

3.4 策略收益

收益是博弈模型中的一个重要元素, 它直接说明了策略的好与坏. 在计算策略收益前, 先描述一下本文模型的交互场景:

1) 一次机会. 本文将协作系统建模成一个时间离散型模型, 在一个时间步里, 每一个用户平均意义上都有一次机会发起一个服务请求.

2) 混合均匀的网络结构 (Well-mixed topology). 此时, 每一个用户都是互相连接的, 且能够以相同的概率进行交互.

3) 一个请求者与一个提供者. 本文假设每一个服务请求者会选择其他一个用户作为其服务提供者. 比如, 在百度云文件共享平台里, 文件下载者会从一个云用户处获得完整的资源. 单请求者多提供者模型可以参考其他文献[4, 14].

本文使用 $N_i(t)$ 表示 t 时刻 s_i 策略在整个系统中的规模, 而 s_i 种群在系统中的比例可以使用 $x_i(t) = N_i(t)/N$ 进行表示, 并且有 $\sum_{i=1}^{151} x_i(t) = 1$. 为了方便起见, 在后面的章节中, 本文会将 $N_i(t)$ 和 $x_i(t)$ 简化表示为 N_i 和 x_i . 此时, 可引入 $X(t) = [x_1, x_2, x_3]$ 表示 t 时刻系统的状态.

首先计算 ALLC 策略的期望收益.

当 ALLC 用户作为服务请求者时, 在混合均匀网络结构假设下, 他/她会分别以 x_1, x_2, x_3 的概率选择 ALLC, ALLD, R 用户作为其交互对象. 由各个策略的性质可得, ALLC 策略的期望收获可以表示为:

$$\begin{aligned} E_1^R(t) &= \alpha(x_1 \cdot p_{11} + x_2 \cdot p_{21} + x_3 \cdot p_{31}) \\ &= \alpha(x_1 + f x_3) \end{aligned} \quad (1)$$

其中 $\alpha > 0$ 表示获得一个服务时用户的收益.

当 ALLC 用户作为服务提供者时,在一个时间步里,他/她会为 ALLC 和 ALLD 用户分别提供 x_1 和 x_2 次服务. 基于交互历史,推荐系统^[4]会将 N_1 个 ALLC 用户和 N_3 个 R 用户判定为合作性参与者,因为 R 用户会从合作性用户那里请求服务,所以平均意义上,每个 ALLC 服务提供者会以 $x_3/(x_1+x_3)$ 的次数为 R 用户提供服务. 综上,ALLC 策略的期望损失为:

$$\begin{aligned} E_1^p(t) &= \beta(x_1 \cdot p_{11} + x_2 \cdot p_{12} + \frac{x_3}{x_1+x_3} \cdot p_{13}) \\ &= \beta(x_1 + x_2 + \frac{x_3}{x_1+x_3}) \end{aligned} \quad (2)$$

其中 $\beta > 0$ 表示提供一次服务时用户的损失.

总而言之, t 时刻 ALLC 策略的期望收益可以表示为:

$$\begin{aligned} \bar{U}_1(t) &= E_1^R(t) - E_1^p(t) \\ &= \alpha(x_1 + fx_3) - \beta(x_1 + x_2 + \frac{x_3}{x_1+x_3}) \end{aligned} \quad (3)$$

同理,可以计算出 t 时刻 ALLD 和 R 策略的期望收益为:

$$\bar{U}_2(t) = \alpha x_1 \quad (4)$$

$$\bar{U}_3(t) = \alpha - \beta(fx_1 + \frac{x_3}{x_1+x_3}) - \gamma \quad (5)$$

其中 $\gamma > 0$ 表示用户采用 R 策略时需要为推荐系统支付的额外花费.

随后,通过公式(6),可以计算出 t 时刻系统的平均收益:

$$\begin{aligned} \bar{U}(t) &= x_1 \cdot \bar{U}_1(t) + x_2 \cdot \bar{U}_2(t) + x_3 \cdot \bar{U}_3(t) \\ &= (\alpha - \beta)(x_1 + x_3) - (\alpha - \beta)(1-f)x_1x_3 \end{aligned} \quad (6)$$

在本文模型中, α, β 和 γ 需要满足下面的关系: 首先, $\alpha > \beta$ 需要成立,即获得服务的收益要大于提供服务的损失,否则,合作将无法产生;其次, $\gamma < \alpha - \beta$ 需要成立,否则,推荐花费过高将导致 R 策略无法在系统中存活.

3.5 演化过程

基于策略的收益,本文中引入了两种学习机制:全局平均学习(Global Mean Learning Mechanism, GMLM)^[24]和当前最优学习(Current Best Learning Mechanism, CBLM)^[24]来描述种群的演化过程.

3.5.1 GMLM 学习机制

在交互完成后,用户会随机选择一个邻居,并与其进行收益的比较,如果自己的收益高于所选邻居的收益,则保持当前策略不变,反之,将自己的策略改变成邻居的策略,这就是 GMLM 学习机制. Börgers 等人^[15]证明在 GMLM 学习机制下,种群的演化过程可以使用复制动力学方程(Replicator equations)^[24,12-14]进行表示,于是本文模型中, t 时刻后 ALLC, ALLD 和 R 策略比例的变化可以描述为:

$$\begin{cases} \dot{x}_1(t) = \lambda x_1(t) (\bar{U}_1(t) - \bar{U}(t)) \\ \dot{x}_2(t) = \lambda x_2(t) (\bar{U}_2(t) - \bar{U}(t)) \\ \dot{x}_3(t) = \lambda x_3(t) (\bar{U}_3(t) - \bar{U}(t)) \end{cases} \quad (7)$$

其中, λ 表示策略的学习率,而且各个策略收益与平均收益的差可以具体由公式(8)~公式(10)表示:

$$\begin{aligned} \bar{U}_1(t) - \bar{U}(t) &= -\alpha(1-f)(x_3 - x_1x_3) + (\beta + \gamma)x_3 - \\ &\quad \beta(x_2 + \frac{x_3}{x_1+x_3}) - \beta(1-f)x_1x_3 \end{aligned} \quad (8)$$

$$\begin{aligned} \bar{U}_2(t) - \bar{U}(t) &= \beta x_1 + (\beta + \gamma - \alpha)x_3 + \\ &\quad (\alpha - \beta)(1-f)x_1x_3 \end{aligned} \quad (9)$$

$$\begin{aligned} \bar{U}_3(t) - \bar{U}(t) &= \alpha x_2 - \beta \frac{x_3}{x_1+x_3} - \gamma + (\gamma + \beta)x_3 + \\ &\quad (1-f)x_1(\beta + (\alpha - \beta)x_3) \end{aligned} \quad (10)$$

定义 1. 在收益指导下,如果 s_i 策略能够最终被所有理性用户选择,即 $x_i = 1$,那么可称 s_i 策略占优了整个系统.

推论 1. 在 GMLM 下,ALLC 策略不能占优整个系统.

证明:这里使用反证法.假设 ALLC 策略能够占优整个系统,那么在 $x_1 \rightarrow 1, x_2 \rightarrow 0, x_3 \rightarrow 0$ 时,下列 3 个不等式将会成立:

$$\bar{U}_1(t) - \bar{U}(t) > 0$$

$$\bar{U}_2(t) - \bar{U}(t) < 0$$

$$\bar{U}_3(t) - \bar{U}(t) < 0$$

由 $\bar{U}_2(t) - \bar{U}(t) < 0$ 可推出 $\beta < 0$,这与本文 $\beta > 0$ 的设定矛盾.所以,当 $\beta > 0$ 时,ALLC 策略将不能占优整个系统.

推论 2. 在 GMLM 下,ALLD 策略不能占优整个系统.

证明:这里使用反证法.假设 ALLD 策略能够占优整个系统,那么在 $x_1 \rightarrow 0, x_2 \rightarrow 1, x_3 \rightarrow 0$ 时,下列 3 个不等式将会成立:

$$\bar{U}_1(t) - \bar{U}(t) < 0$$

$$\bar{U}_2(t) - \bar{U}(t) > 0$$

$$\bar{U}_3(t) - \bar{U}(t) < 0$$

由 $\bar{U}_3(t) - \bar{U}(t) < 0$ 可推出 $\alpha x_2 - \beta x_3/(x_1+x_3) - \gamma < 0$. 因为 $x_3/(x_1+x_3) > 0$,所以 $\alpha x_2 - \gamma < 0$ 是 $\bar{U}_3(t) - \bar{U}(t) < 0$ 成立的充分条件.因此有 $\alpha < \gamma$,这与 $\alpha > \gamma$ 的设定矛盾.所以,当 $\alpha > \gamma$ 时,ALLD 策略将不能占优整个系统.

推论 3. 在 GMLM 下,当 $f \geq 1 - \gamma/\alpha$ 时,R 策略不能占优整个系统.

证明:这里使用反证法.假设当 $f \geq 1 - \gamma/\alpha$ 时,R 策略能够占优整个系统.那么在 $x_1 \rightarrow 0, x_2 \rightarrow 0, x_3 \rightarrow 1$ 时,下列 3 个不等式将会成立:

$$\bar{U}_1(t) - \bar{U}(t) < 0$$

$$\bar{U}_2(t) - \bar{U}(t) < 0$$

$$\bar{U}_3(t) - \bar{U}(t) > 0$$

由 $\bar{U}_1(t) - \bar{U}(t) < 0$ 可推出 $-\alpha(1-f) + \gamma < 0$,进而有 $f < 1 - \gamma/\alpha$,这与 $f \geq 1 - \gamma/\alpha$ 的假设矛盾,所以 $f \geq 1 - \gamma/\alpha$ 时,R 策略将不能占优整个系统.

下面本文将讨论两种特殊的情况: $f=1$ 和 $f=0$.

推论 4. 在 GMLM 下,当 $f=1$ 时,那么 ALLC, ALLD 和 R 策略都不能占优整个系统.

证明:参见推论 1、推论 2 和推论 3.

推论 5. 在 GMLM 下,当 $f=0$ 时,那么 R 策略会占优整个系统.

证明:此时可以计算 ALLD 策略与 ALLC 策略的收益差为:

$$\bar{U}_2(t) - \bar{U}_1(t) = \beta(x_1 + x_2 + \frac{x_3}{x_1+x_3}) > 0$$

所以,与 ALLD 策略相比,ALLC 策略是一个严格的劣策略,根据博弈论中重复剔除劣策略方法,可将 ALLC 从策略集中删除,即令 $x_1 = 0$. 然后,可以计算 R 策略与 ALLD 策略的收益差为:

$$\bar{U}_3(t) - \bar{U}_2(t) = \alpha - \beta - \gamma$$

所以,在 $\gamma < \alpha - \beta$ 的设定下,有 $\bar{U}_3(t) - \bar{U}_2(t) > 0$,即,与

R 策略相比, ALLD 策略成为了一个严格劣策略, 再一次使用劣策略剔除法, ALLD 可以从策略集中删除. 此时, 系统中就只剩下 R 策略了, 即 R 策略占优了整个系统.

3.5.2 CBLM 学习机制

在交互完成后, 设 t 时刻系统里期望收益最高的策略为 s_h , 其对应的收益为 $\bar{U}_h(t)$. 用户通过获取邻居的收益会感知出当前系统中期望收益最高的策略 s_h , 如果自己的策略与 s_h 相同, 则保持不变, 否则就把自己的策略改变成 s_h , 这就是 CBLM 学习机制. 并且可以使用公式(11)形式化表示 CBLM 机制下群体的演化:

$$\begin{cases} \dot{x}_i(t) = \lambda x_i(t) (\bar{U}_i(t) - \bar{U}_h(t)) (i \neq h) \\ \dot{x}_h(t) = \lambda \sum_{i=1, i \neq h}^{|S|} x_i(t) (\bar{U}_h(t) - \bar{U}_i(t)) \end{cases} \quad (11)$$

为了方便起见, 在讨论模型在 CBLM 机制下的演化特性时, 本文这里只考虑了 $f=0$ 时的场景, 此时有:

$$\begin{cases} \bar{U}_1(t) = \alpha x_1 - \beta(x_1 + x_2 + \frac{x_3}{x_1 + x_3}) \\ \bar{U}_2(t) = \alpha x_1 \\ \bar{U}_3(t) = \alpha - \beta \frac{x_3}{x_1 + x_3} - \gamma \end{cases} \quad (12)$$

推论 6. 在 CBLM 下, ALLC 策略不能占优整个系统.

证明: 这里使用反证法. 假设在 CBLM 下, ALLC 策略能够占优系统, 那么有:

$$\begin{aligned} \bar{U}_1(t) &> \bar{U}_2(t) \\ \bar{U}_1(t) &> \bar{U}_3(t) \end{aligned}$$

显然 $\bar{U}_1(t) > \bar{U}_2(t)$ 在 $\beta > 0$ 的设定下是不会成立的, 这与假设矛盾, 从而推论得证.

推论 7. 在 CBLM 下, ALLD 策略不能占优整个系统.

证明: 这里使用反证法. 假设在 CBLM 下, ALLD 策略能够占优系统, 那么有:

$$\begin{aligned} \bar{U}_2(t) &> \bar{U}_1(t) \\ \bar{U}_2(t) &> \bar{U}_3(t) \end{aligned}$$

其中, $\bar{U}_2(t) - \bar{U}_3(t) = -\alpha(x_2 + x_3) + \beta \frac{x_3}{x_1 + x_3} + \gamma \leq -\alpha(x_2 + x_3) + \beta + \gamma$, 随着 x_2 的增加, x_1 与 x_3 的减少, $-\alpha(x_2 + x_3)$ 会持续的减小, 在 $\gamma < \alpha - \beta$ 的设定下, 总会存在一个时间 t , 使得 $-\alpha(x_2 + x_3) + \beta + \gamma < 0$ 成立, 进而有: $\bar{U}_2(t) < \bar{U}_3(t)$, 这与 $\bar{U}_2(t) > \bar{U}_3(t)$ 的假设矛盾, 所以在 CBLM 下, ALLD 策略不能占优整个系统.

推论 8. 在 CBLM 下, 当 $x_2 + x_3 > (\beta + \gamma)/\alpha$ 时, R 策略能够占优整个系统.

证明: 由于 $\bar{U}_2(t) > \bar{U}_1(t)$ 在 $f=0$ 时会恒成立, 所以只需要考虑 $\bar{U}_3(t) > \bar{U}_2(t)$ 的可行性即可. 首先计算:

$$\begin{aligned} \bar{U}_3(t) - \bar{U}_2(t) &= \alpha(x_2 + x_3) - \beta \frac{x_3}{x_1 + x_3} - \gamma \\ &\geq \alpha(x_2 + x_3) - \beta - \gamma \end{aligned}$$

当 $x_2 + x_3 > (\beta + \gamma)/\alpha$ 时, 可以推出 $\bar{U}_3(t) - \bar{U}_2(t) > 0$, 此时 R 策略就是当前最优策略. 由于下一时间步, $x_2 + x_3$ 仍将保持增长, 所以 $x_2 + x_3 > (\beta + \gamma)/\alpha$ 的条件会保持不变, 进而, 可以推出 $\bar{U}_3(t) > \bar{U}_2(t) > \bar{U}_1(t)$ 在未来的种群演化过程中保持不变, 直到 R 策略占优整个系统. 另一点需要说明的是, 在本文模型中 $x_2 + x_3 > (\beta + \gamma)/\alpha$ 是必然会发生的. 那是

因为在 $f=0$ 和 CBLM 设定下, 由于 $\bar{U}_1(t) < \bar{U}_2(t)$ 恒成立, 所以 ALLC 策略的比例会一直降低, 当 x_1 降低到 $1 - (\beta + \gamma)/\alpha$ 时, R 策略占优系统的条件也就成熟了.

4 一些实际问题

4.1 非完美推荐

第 3 章在构建模型时只考虑了完美推荐机制场景, 即推荐系统能够准确地分辨出用户的合作属性, 进而 R 策略用户在每次交互中都能成功地获得服务. 然而, 由于数据获取时的噪声以及数据稀疏性等因素的影响, 推荐系统往往不能做到 100% 的准确, 因此, 为了符合实际场景, 还需要考虑在非完美推荐下本文模型的表现.

在非完美推荐下, 系统可能会发生两种类型的错误: 其一, 是将背叛性用户误判成合作性用户; 其二, 是将合作性用户误判成背叛性用户. 由于在现实场景里, 背叛性用户会使用共谋、洗白等方法以提高自己被误判成合作性用户的概率, 进而可以获得更多的服务, 因此讨论理性背叛模型在第一种推荐错误下的性能是十分有必要的事情.

为了刻画第一类错误, 本文引入参数 $\delta \in [0, 1]$ 表示 ALLD 用户被误认为是合作性用户的概率. 那么此时, R 策略用户选取交互对象的规模就从 $N_1 + N_3$ 扩大为了 $N_1 + \delta N_2 + N_3$.

接下来, 计算在第一类错误模型下各个策略的期望收益. 首先是 ALLC 策略的期望收益.

当 ALLC 用户作为服务请求者时, 他/她仍然能以 100% 和 f 的概率从 ALLC 和 R 交互对象中获得服务, 故其期望收益保持不变, 如公式(1)所示. 当 ALLC 用户作为服务提供者时, 其被 R 用户访问到的次数将变化为:

$$\frac{N_3}{N_1 + \delta N_2 + N_3} = \frac{x_3}{x_1 + \delta x_2 + x_3}$$

与之相对应地, ALLC 策略的期望损失更新为:

$$E_1^P(t) = \beta(x_1 + x_2 + \frac{x_3}{x_1 + \delta x_2 + x_3})$$

综上, 在第一类错误模型下, ALLC 策略的期望收益为:

$$\bar{U}_1^{First}(t) = \alpha(x_1 + f x_3) - \beta(x_1 + x_2 + \frac{x_3}{x_1 + \delta x_2 + x_3}) \quad (13)$$

在计算 ALLD 策略期望收益时, 此处考虑了最坏的情况, 即 ALLD 策略会以 δ 的概率被误判为 R 策略, 此时 ALLD 策略的期望收益可以表示为:

$$\bar{U}_2^{First}(t) = \alpha(x_1 + \delta x_3) \quad (14)$$

由于第一类错误的存在, 以 $(x_1 + x_3)/(x_1 + \delta x_2 + x_3)$ 的概率, R 用户会获得服务, 以 $f x_1 + \delta x_2 + x_3/(x_1 + \delta x_2 + x_3)$ 的次数, R 用户给其他用户提供服务, 因此, R 策略的期望收益可以表示为:

$$\bar{U}_3^{First}(t) = \alpha \frac{x_1 + x_3}{x_1 + \delta x_2 + x_3} - \beta(f x_1 + \delta x_2 + \frac{x_3}{x_1 + \delta x_2 + x_3}) - \gamma \quad (15)$$

在非完美推荐场景下, 模型的鲁棒性是一个值得关注的问题. 具体而言, 本文希望可以找到错误概率 δ 的临界值 δ_c , 当错误概率高于该临界值时, ALLD 策略会占优整个系统. 笔者之前的工作^[4]讨论了 $f=1$ 时, δ_c 的近似解. 本文将进一步讨论 $f=0$ 时, δ_c 的求解.

当 $f=0$ 时,各个策略的期望收益如下所示:

$$\begin{cases} \bar{U}_1^{First}(t) = \alpha x_1 - \beta(x_1 + x_2 + \frac{x_3}{x_1 + \delta x_2 + x_3}) \\ \bar{U}_2^{First}(t) = \alpha(x_1 + \delta x_3) \\ \bar{U}_3^{First}(t) = \alpha \frac{x_1 + x_3}{x_1 + \delta x_2 + x_3} - \beta(\delta x_2 + \frac{x_3}{x_1 + \delta x_2 + x_3}) - \gamma \end{cases}$$

ALLD 策略占优整个系统的充要条件为:

$$\begin{cases} \bar{U}_2^{First}(t) - \bar{U}_1^{First}(t) > 0 \\ \bar{U}_2^{First}(t) - \bar{U}_3^{First}(t) > 0 \end{cases}$$

首先易知, $\bar{U}_2^{First}(t) - \bar{U}_1^{First}(t) > 0$ 在 $f=0$ 设定下是成立的.

基于 $\bar{U}_2^{First}(t) > \bar{U}_3^{First}(t)$ 可推出:

$$\delta > 1 - \beta \frac{\frac{x_3}{x_1 + \delta x_2 + x_3} + \delta x_2}{\alpha x_3} - \frac{\gamma}{\alpha x_3}$$

由于:

$$\text{MIN} \left(\beta \frac{\frac{x_3}{x_1 + \delta x_2 + x_3} + \delta x_2}{\alpha x_3} + \frac{\gamma}{\alpha x_3} \right) = \frac{\beta + \gamma}{\alpha}$$

此时有: $\delta \geq 1 - (\beta + \gamma)/\alpha$ 时 ALLD 策略能够占优整个系统. 故 $\delta_c = 1 - (\beta + \gamma)/\alpha$. 注意该临界值是一个近似解,它表示了 ALLD 策略能够占优整个系统的充分条件.

4.2 ALLC 与 R 策略的识别

如果每个用户都如实地公布自己所选择的策略,那么服务提供者就可以很清楚地根据对手类型,从而做出相应的反馈.然而,在更多的实际场景里,用户往往是不会公布自己的策略,此时,就需要根据用户间的交互历史来推断用户的类型.在之前的文献[4]里,笔者基于 Eigentrust 算法实现了用户合作度的评估,然而,这种方法不能直接用作 ALLC 和 R 策略的识别,因为他们的合作度可能都会很高.为了解决这一问题,本文提出了一种基于相似度的策略识别法.

首先引入 R_{ij} 和 P_{ij} 分别表示当前时间步 t 里用户 i 接收到的来自于用户 j 的服务请求数量和用户 i 为用户 j 提供服务的数量.那么用户 i 对用户 j 的合作度可以表示为:

$$c_{ij} = \begin{cases} \frac{P_{ij}}{R_{ij}} & \text{if } R_{ij} > 0 \\ 0.5 & \text{otherwise} \end{cases} \quad (16)$$

基于此,可以构建出一个衡量用户间合作度关系的 $N \times N$ 维的矩阵 $C = [c_{ij}]$.由于 ALLC 是一个无条件合作者,他/她会为任何用户都提供服务,而 R 策略是不会为 ALLD 策略提供服务的,因此,可以通过衡量用户之间的行为相似度来帮助 R 策略使用者进行策略识别.

本文采用余弦相似度来计算任意两个用户 i 和 j 之间的行为相似度 $\text{sim}(i, j)$:

$$\text{sim}(i, j) = \begin{cases} \frac{\sum_{k \in \text{comm}(i, j)} c_{ik} \cdot c_{jk}}{\sqrt{\sum_{k \in \text{comm}(i, j)} (c_{ik})^2} \cdot \sqrt{\sum_{k \in \text{comm}(i, j)} (c_{jk})^2}} & \text{if } |\text{comm}(i, j)| > 0 \\ 0.5 & \text{otherwise} \end{cases} \quad (17)$$

其中, $\text{comm}(i, j)$ 表示共同向用户 i 和 j 发起过服务请求

的用户集合.随后,本文引入一个阈值 θ ,当两个用户的行为相似度大于 θ 时,可认为二者采用了相同的策略,否则,可认为二者采用了不同的策略.综上,在本文提出的理性背叛模型中, R 策略的行为可以由算法 1 表示.

算法 1. 基于相似度识别法的 R 策略行为决策

输入:

i : R 策略的服务提供者

j : 服务请求者

输出:

p : i 对 j 提供服务的概率

1. 基于文献[4],对 j 的合作度进行衡量

2. if j 是一个背叛性用户

3. $p = 0$

4. else

5. 使用公式(17)计算 i 与 j 的相似度 $\text{sim}(i, j)$

6. if $\text{sim}(i, j) > \theta$

7. $p = 1$

8. else

9. $p = f$

10. return p

5 实验及分析

本节将使用数值实验和仿真实验说明本文理性背叛模型的有效性,并对本文的理论分析进行了验证.

5.1 参数设置

实验中的参数设置如表 1 所示.在本文模型里,获得一个服务的收益 α 大于提供一次服务的损失 β ,为了和文献[24, 14]保持一致,实验取 $\alpha = 7$ 和 $\beta = 1$. R 策略采取推荐服务时需要付出一定的花费 γ ,为了与笔者之前的文献[4]保持一致,实验取 $\gamma = 0.7$.在本文提出的理性背叛模型中, R 策略与 ALLC 策略进行合作的概率为 f ,该值是一个从 0 到 1 的连续值.因为在现实场景里,只有很小部分的用户会及时去感知收益并进行策略更新,因此实验设置学习率 $\lambda = 0.01$.在仿真实验中,系统中的用户数量被设置为 1000.在基于相似度的策略识别法里,实验将阈值 θ 分别设置为 0.4, 0.5, 0.6 和 0.7.

表 1 参数设置
Table 1 Parameters setting

参数	定义	值
α	获得一个服务的收益	7
β	提供一个服务的损失	1
γ	采用推荐服务的花费	0.7
f	R 给 ALLC 提供服务的概率	[0, 1]
λ	学习率	0.01
N	系统中的用户数量	1000
θ	相似度阈值	0.4, 0.5, 0.6, 0.7

5.2 f 对模型的影响

首先进行数值实验研究在完美推荐场景下 f 对模型的影响.图 1 展示了 GMLM 学习机制下不同 f 值对应的策略演化结果.从图 1(a)和图 1(b)中可知,当 $f=0$ 和 $f=0.8$ 时, R 策略都会占优整个系统,使 ALLD 策略从系统中消失.而且,相比 $f=0.8$ 场景, $f=0$ 时 R 策略可以获得更好地收益,此时 R

策略可以更快地占优整个系统.如图1(c)和图1(d)所示,当 $f \geq 0.9$ 时,ALLC,ALLD,R会互相占优并共存于系统之中.图1的实验结果很好地验证了本文3.5节中提出的推论1至推论5.当 f 取其他值时,策略演化结果是相似的,所以本文中就不进行展示了.

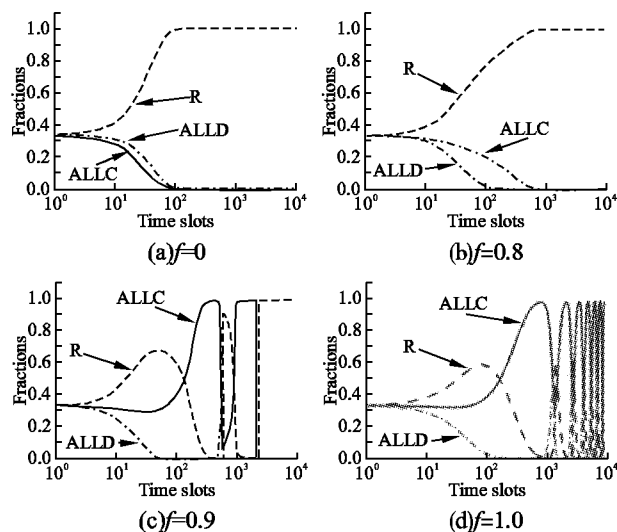


图1 GMLM学习机制下 f 对模型的影响[数值实验]

Fig. 1 Impact of f on the model with GMLM
[numerical experiments]

图2展示了CBLM学习机制下不同 f 值对应的策略演化结果.与图1(a)和图1(b)相似,如图2(a)和图2(b)所示,在 $f=0$ 和 $f=0.8$ 设置下,R策略可以占优整个系统,并且 f 值越小,R策略占优系统所花费的时间就越少.图2(a)所展示的

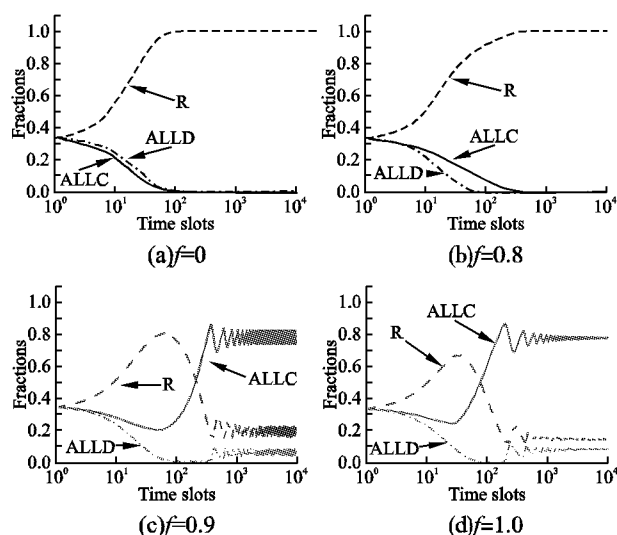


图2 CBLM学习机制下 f 对模型的影响[数值实验]

Fig. 2 Impact of f on the model with CBLM
[numerical experiments]

结果验证了本文3.5节中提出的推论6至推论8.与GMLM机制下的策略演化不同,如图2(c)和图2(d)所示,当 $f=0.9$ 和 $f=1.0$ 时,ALLC,ALLD,R不仅会共存于系统之中,而且它们的比例会趋向于收敛.在之前的文献[4]中,笔者已经详

细地解释了 $f=1.0$ 时系统会进行收敛的原因,此处就不赘述了.同样地,当 f 取其他值时,策略演化会呈现相似的结果,所以本文就不进行展示了.

基于图1和图2两个实验,可以发现:与不考虑理性背叛的模型($f=1.0$)相比,考虑了理性背叛的模型($f<1$)可以进一步促进合作的涌现,而且在本文参数设定下,当 $f<0.9$ 时,无论在GMLM还是CBLM学习机制下,ALLD策略都会从系统中消失,从而合作行为获得了极大的激励.

为了验证数值实验的准确性,本文还进行了相应的基于智能体的仿真实验.图3和图4分别展示了用户数量

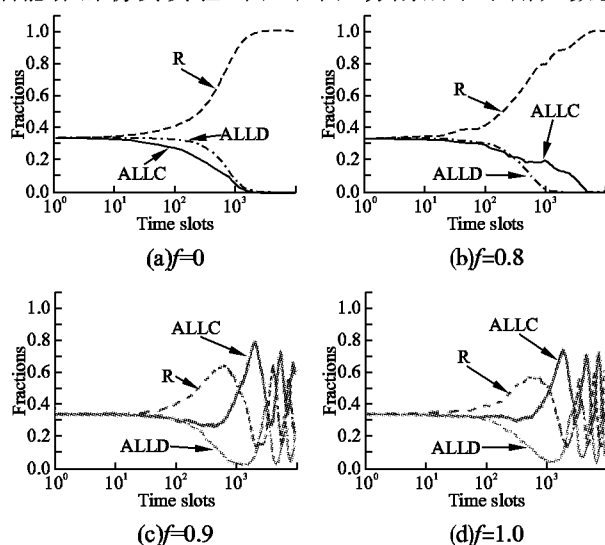


图3 GMLM学习机制下 f 对模型的影响[仿真实验]

Fig. 3 Impact of f on the model with GMLM
[simulation experiments]

$N=1000$ 时,在GMLM和CBLM学习机制下,系统中策略的演化过程.通过对比可知,仿真实验取得了与数值实验一致的结果.

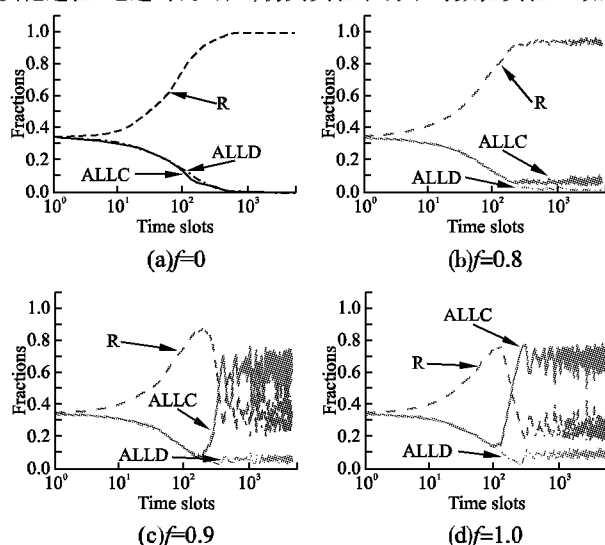


图4 CBLM学习机制下 f 对模型的影响[仿真实验]

Fig. 4 Impact of f on the model with CBLM
[simulation experiments]

结果.当 $N=5000, 10000$ 时,仿真实验也取得了相似的结果,

本文就不进行展示了。

5.3 非完美推荐

本节将讨论在非完美推荐场景下,本文模型的性能。由于仿真实验与数值实验具有相似的实验结果,为了节约篇幅,这里主要以数值实验为例说明在 $f=0$ 时,错误概率 δ 对策略演化的影响。图 5 展示了在 GMLM 学习机制下系统稳定时策略比例的状态,可以发现错误概率的临界值 $\delta_c = 0.7$,当 δ 低于该值时,R 策略可以占优整个系统,当 δ 高于该值时,ALLD 策略会占优整个系统。需要注意的是,在 $f=0$ 时,ALLC 是一个严格劣策略,系统中不会有人采用该策略。而且,在 4.1 节中,近似计算的错误临界值 $\delta_c = 1 - \frac{1.7}{7} = 0.757$,已经十分接近了实验中得到的临界值。在 CBLM 学习机制下,也可得到相似的结果,此处就不进行赘述了。而且,在笔者之前文献[4]构建的合作推荐系统里,推荐错误率大约在 10% 左右,此时若采用理性背叛模型,R 策略可以占优整个系统。

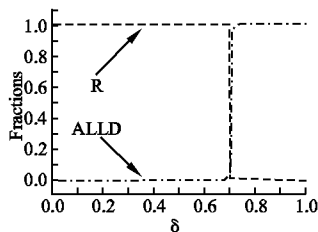


图 5 GMLM 机制下的非完美推荐模型

Fig. 5 Imperfect recommendation model with CBLM

5.4 相似度策略识别

这里,通过仿真实验研究了本文机制在部署相似度策略识别法后的策略演化过程。如图 6(a) 至图 6(d) 所示,实验分

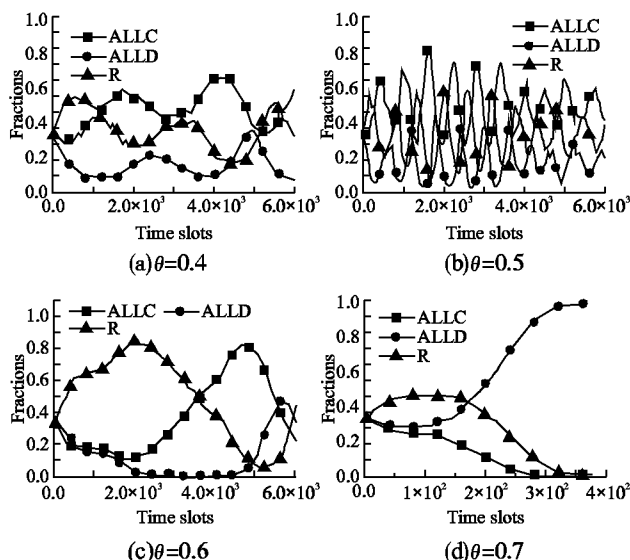


图 6 基于余弦相似度的策略识别

Fig. 6 Strategy recognition based on cosine similarity

别设置相似度阈值 $\theta = 0.4, 0.5, 0.6$ 和 0.7 。此处设定用户采用 GMLM 作为其策略学习机制。从图 6(a) 和图 6(b) 中可以发现,如果把阈值 θ 设置为一个较小的值时,R 策略会大概率地将 ALLC 策略判断成 R 策略(此时等价于将 f 设置成了一

个较大的值),并为他们提供服务,从而出现了如图 1(c) 和图 1(d) 所示的策略比例反复振荡的现象。图 6(c) 展示 $\theta = 0.6$ 时的策略演化过程,在这种设定下,R 策略可以较成功地进行策略识别,所以刚开始时 R 策略的比例可以大幅度地增长,随后,伴着 ALLD 策略的降低,R 策略与 ALLC 策略的相似度会趋近于一致,因此 R 策略会大概率地将 ALLC 识别为 R 策略,进而促进了 ALLC 策略的增长。当 ALLC 策略增长到一定程度后,ALLD 策略会得到激励。最终也会出现策略比例反复震荡的现象。从图 6(d) 中可以发现,如果把阈值 θ 设置为一个较大的值时,R 策略会出现自相伤害的情况,进而不利于合作的产生。

综上,笔者认为 θ 取 0.6 是一个比较好的选择,在该设定下 ALLD 策略大部分时间内都可以被很好地抑制。在未来的工作里,笔者在计算相似度时将引入时间等因素,希望可以进一步提高策略识别的准确性。

5.5 对比实验

这里,通过对比实验将本文提出的具有理性背叛属性的推荐机制策略 R 与下列六种激励策略进行了性能比较: RIM^[4], TFT^[5], TF2T^[6], WLSLS^[8], Mirror^[9], Proportion^[2]。在比实验时,本文考虑了 4500 个用户,每个策略初始化时都拥有 500 个用户作为其策略种群。不失一般性,此处将 R 策略的合作参数 f 设置为 0。如图 7 所示,随着系统的演化,R 策略会逐渐战胜其他六种激励策略,并最终占优整个系统。在本文模型中,基于推荐机制,R 策略能够以更高的概率获得服务,而且由于理性背叛机制的引入,R 策略也避免了被 ALLC 策略所抑制,从而 R 策略可以在多策略对比实验中获得不错的表现。

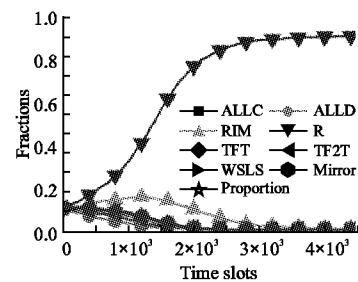


图 7 多策略模型对比实验

Fig. 7 Performance comparison in a multi-strategies model

6 总结与展望

为了解决无条件合作策略对合作涌现的抑制现象,本文引入了理性背叛模型。在演化博弈框架下,针对两种不同的用户学习机制,本文研究了背叛概率对种群演化的影响,理论证明理性背叛机制的引入可以更好地促进合作的产生。接着,针对实际场景,本文首先研究了非完美推荐下模型的鲁棒性问题,然后,为了进行策略区分,进一步提出了基于相似度的策略识别法。数值实验和仿真实验都证明了理性背叛机制在合作促进方面的有效性能。在部署相似度策略识别法时,发现在动态演化的环境下准确地判断用户类型不是有一件容易的事情,因此,笔者将会继续对该问题进行研究。

References:

- [1] Hughes Daniel, Coulson Geoff, Walkerdine James. Free riding on Gnutella revisited; the bell tolls? [J]. IEEE Distributed Systems Online, 2005, 6(8): 1-18.
- [2] Zhao B Q, Lui J C S, Chin D M. A mathematical framework for analyzing adaptive incentive protocols in P2P networks [J]. IEEE/ACM Transactions on Networking (TON), 2012, 20(2): 367-380.
- [3] Wang Yu-feng, Nakao Akihiro, Vasilakos Athanasios V, et al. P2P soft security: on evolutionary dynamics of P2P incentive mechanism [J]. Computer Communications, 2011, 34(3): 241-249.
- [4] Li Ming-chu, Jin Xing, Guo Cheng, et al. RIMNet: recommendation incentive mechanism based on evolutionary game dynamics in peer-to-peer service networks [J]. Knowledge-Based Systems (KBS), 2019, 166: 156-169.
- [5] Axelrod Robert, Hamilton William D. The evolution of cooperation [J]. Science, 1981, 211(4489): 1390-1396.
- [6] Axelrod Robert, Dion, Douglas. The further evolution of cooperation [J]. Science, 1988, 242(4884): 1385-1390.
- [7] Nowak Martin A, Sigmund Karl. Tit for tat in heterogenous populations [J]. Nature, 1992, 355(6357): 250-253.
- [8] Nowak Martin A, Sigmund Karl. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game [J]. Nature, 1993, 364(6432): 56-58.
- [9] Feldman M, Chuang J. Overcoming free-riding behavior in peer-to-peer systems [J]. ACM Special Interest Group on Economics and Computation (SIGECOM), 2005, 5(4): 41-50.
- [10] Nowak Martin A, May Robert M. Evolutionary games and spatial chaos [J]. Nature, 1992, 359(6398): 826-829.
- [11] Santos Francisco C, Pacheco Jorge M. Scale-free networks provide a unifying framework for the emergence of cooperation [J]. Physical Review Letters, 2005, 95(9): 1-4.
- [12] Tian Lin-lin, Li Ming-chu, Wang Zhen. Cooperation enhanced by indirect reciprocity in spatial prisoner's dilemma games for social P2P systems [J]. Physica A: Statistical Mechanics and its Applications, 2016, 462: 1252-1260.
- [13] Smith John Maynard. Evolution and the theory of games [M]. Cambridge: Cambridge University Press, 1982.
- [14] Cui Guang-hai, Li Ming-chu, Jin Xing, et al. Evolution of cooperation in P2P media sharing networks based on self-organized incentive [J]. Journal of Chinese Computer Systems, 2016, 37(2): 202-206.
- [15] Börgers Tilman, Rajiv Sarin. Learning through reinforcement and replicator dynamics [J]. Journal of Economic Theory, 1997, 77(1): 1-14.

附中文参考文献:

- [14] 崔光海, 李明楚, 金星, 等. 基于自组织激励的 P2P 多媒体共享网络节点合作的演化 [J]. 小型微型计算机系统, 2016, 37(2): 202-206.

~~~~~

## 《小型微型计算机系统》征订启事

《小型微型计算机系统》创刊于 1980 年,由中国科学院主管,中国科学院沈阳计算技术研究所主办,中国计算机学会会刊(月刊),国内外公开发行。

《小型微型计算机系统》内容涵盖计算机学科各领域,包括:计算机科学理论、体系结构、数据库理论、计算机网络与信息安全、人工智能与算法、服务计算、计算机图形与图像等。

收录情况:中文核心期刊;中国学术期刊文摘(中英文版);中国科学引文数据库(CSCD)来源期刊;英国《科学文摘》(INSPEC);美国《剑桥科学文摘(自然科学)》(CSA(NS)); Cambridge Scientific Abstracts(Natural Science)等。

国内邮发代号:8-108

国外发行代号:M349

国内统一连续出版物号:CN21-1106/TP

国际标准连续出版物号:ISSN1000-1220

编辑部地址:沈阳市东陵区南屏东路 16 号《小型微型计算机系统》编辑部

邮政编码:110168

电话:(024)24696120

E-mail:xwjxt@sict.ac.cn

网址:http://xwxt.sict.ac.cn