参赛学生姓名：陈昊宇，齐奕安

中学：北京十一学校

省份：北京

国家/地区：中国

指导老师姓名：王艺臻

指导老师单位：北京十一学校

论文题目：多轮博弈中信用系统对于打破囚徒困境的作用

多轮博弈中信用系统对于打破囚徒困境的作用

陈昊宇，齐奕安

摘要

囚徒困境揭示了个体最优策略与全局最优策略之间的矛盾。在理性决策的情况下，整体效益往往较低。为了优化整体效益，需要引入额外信息，以提升博弈双方的合作可能性。本文通过数学建模构建了一个合理的信用体系，使得博弈双方能够通过计算期望收益进行决策。在该信用系统的约束下，背叛不再是理性个体的主导策略，从而使合作在博弈中成为可能。该信用系统的应用范围包括课堂小组作业、竞赛组队等场景。

关键词：囚徒困境，信用系统，期望收益

目录

1 问题背景

在二十世纪五十年代，梅里尔 · 弗勒德和梅尔文 · 德雷希尔提出了囚徒困境 (Tucker, 1950)，并成为了博弈论中的经典问题。具体描述如下：两个罪犯（分别称为 A 和 B）因一起犯罪而被捕，被关押在不同的牢房里，彼此之间没有任何交流。在单独审讯期间罪犯 A 被告知如果他供认并同意指证罪犯 B，他将获得缓刑，罪犯 B 将被监禁 10 年。然而，如果与此同时罪犯 B 认罪并同意指证罪犯 A，罪犯 A 的证词将被撤销，每人将被判处 6 年监禁。罪犯 A 被告知罪犯 B 也得到了同样的待遇。罪犯 A 和罪犯 B 都知道如果双方都不指证对方，他们只能以较轻的罪名被定罪判 3 年监禁。

表 1: 经典囚徒困境

	罪犯 B		
罪犯 A	决策	合作	背叛
	合作	皆被判三年	A 被判十年，B 被判缓刑
	背叛	A 被判缓刑，B 被判十年	皆被判六年

在这个情境中，罪犯之间不能沟通，都会争取最少的刑罚，且没有除描述之外的利益交换。

可知，无论对方怎么选择，罪犯选择背叛的收益均高于选择合作的收益。比如，当罪犯 B 选择合作时，罪犯 A 应该背叛，因为缓刑优于被判三年；若罪犯 B 选择背叛，罪犯 A 仍应选择背叛，因为被判六年优于被判十年。同理，该论断站在罪犯 B 的视角也成立。因此，结果一定是双方背叛，即同时被判六年。该情景下，背叛是双方的主导策略，双方共同背叛被称作做该博弈的纳什均衡 (Kreps, 1989)。

然而，该情景下，纳什均衡不是对于整体最优的结果，但这又是双方理性抉择所必然导致的。当博弈仅进行一次时，双方都背叛是必然发生的；但是当博弈进行多次时，合作有可能产生。过去针对双人博弈的研究已经提到了这一点 (Raihani & Bshary, 2011; Leinfellner, 1986)。而对于多轮次博弈的双人场景中，过往研究聚焦于寻找一个参与者的最佳策略。例如著名的 Tit-for-tat 策略 (Axelrod, 1984)，即永远模仿对手之前的行为，和各种适应性策略 (Hadzikadic & Sun, 2007; Press & Dyson, 2012)。

从组织者的角度出发，囚徒困境阻碍了整体效益的最大化，是应该避免的。因此，本文旨在通过引入一个信任机制以提高整体效益。

2 情景描述

共 N 个个体进行无限轮数的博弈，每一轮中所有个体会被随机配对。个体可以选择合作或背叛他的对手；如果博弈双方均合作，则下一轮不会被分配新的对手，否则在下一轮随机与另一个体配对。

每一轮次的博弈两人根据选择都有如下的收益分布（$r > p > s > q$ 且 $2p > r + q$）

表 2: 囚徒困境下的双方收益矩阵

	个体 2		
个体 1	决策	合作	背叛
	合作	p, p	q, r
	背叛	r, q	s, s

3 引入模型

为了找到适合合作者的环境，该引入模型被提出并建立。

3.1 基本假设

1. 每个个体面对的收益矩阵都是相同的，即环境以同样标准要求所有人。
2. 每个个体的初始信用，即系统对其给出的合作概率为 P_{init}。每当该个体，做出合作或是背叛的决定后，系统会调整其合作概率，$P = \frac{\#合作}{\#决策}$。
3. 收益于一限度 π_{min} 的个体会被强制退出博弈，并补充一个个体，旨在发现该环境下所剩个体的决策倾向，并判断该环境是否适合合作者。

3.2 模型建立

对于个体 i，我们直接建立概率转移矩阵 $\mathbf{T_i}$ 以模拟其决策偏好

$$\mathbf{T_i} = \begin{bmatrix} P_{C|C} & P_{C|D} \\ P_{D|C} & P_{D|D} \end{bmatrix}$$

$P_{C|C}$，$P_{C|D}$，$P_{D|C}$，$P_{D|D}$ 为个体的属性。$P_{C|C}$ 代表对方合作，自身合作的概率，$P_{C|D}$ 代表对方背叛，自身合作的概率，$P_{D|C}$ 代表对方合作，自身背叛的概率 $P_{D|D}$ 代表对方背叛，自身背叛的概率。我们有 $P_{C|C} + P_{D|C} = 1$ 且 $P_{C|D} + P_{D|D} = 1$。

概率转移矩阵的数值某种程度上反应了个体的性格。比如说，$P_{C|C}$ 高就说明了感恩属性，反之 $P_{D|C}$ 高说明见利忘义。而 $P_{C|D}$ 高说明有道德（有道德的判断标准之一便是愿意别人以自己对待其的方式对待自身，而无论如何都选择合作显然是

最为理想的对手），反之则说明是以眼还眼类型的。

同时根据此时的信用系统可以得到个体 j 的预期决策倾向 $\mathbf{E_j}$。

$$\mathbf{E_j} = \begin{bmatrix} P_j \\ 1 - P_j \end{bmatrix}$$

基于个体 i 的概率转移矩阵 $\mathbf{T_i}$，以及个体 j 的预期决策倾向 $\mathbf{E_j}$，可以得到个体 i 面对个体 j 的决策向量 $\mathbf{D_i}$。

$$\mathbf{D_i} = \mathbf{T_i} \cdot \mathbf{E_j}$$

通过调整 $\mathbf{T_i}$ 和 P_{init} 的取值并进行大量模拟，找到了适合合作者的环境。

3.3 引入模型的发现

1. 在现有信用系统下，当 $P_{C|C}, P_{C|D}, P_{D|C}, P_{D|D}$ 随机取值时不论 P_{init} 怎么取值最后经过一定淘汰后剩下的个体都极度倾向于背叛（$P_{D|C}$，$P_{D|D}$ 的值都很大）。

2. 最后个体倾向于合作的情况仅出现在初始时大部分的人都倾向 Tit-for-tat，即 $P_{D|C}$，$P_{D|D}$ 都接近 100% 时。只有这样，才能有效的识别出极度倾向背叛的个体，并将其淘汰，从而维持合作的稳定。

因此，信用系统需要更为严格，并且要鼓励和信誉高的人合作，背叛信誉低的人，这样才可以避免极度倾向背叛的人影响整个环境，从而使得整体保持合作。

4 信用系统

4.1 基本假设

1. 每个个体将会被赋予相同的初始信誉分 c_{init}。

2. 信誉分低于一限度 c_{min} 的个体会被强制退出博弈。

3. 信誉分的增减不仅取决于自身的选择，还取决于对手的守信程度。

4.2 模型建立

本模型的信用系统中分数变化幅度取决于双方的分数，与 elo 等级分系统 (Berg, 2020; Pelánek, 2016) 类似。因此，参考 elo 等级分系统，对个体 i 合作概率的估计为

$$P_i = \frac{1}{1 + \mathrm{e}^{-c_i}}$$

根据基本假设，对于二人博弈，如果遇到了信誉分极低的对象，迫不得已选择背叛是无可厚非的，这种情况下信用系统应酌情降低扣分幅度。同理，与背叛信誉分高的对象应导致很大程度的扣分。记 $A_i \in \{0, 1\}$ 为个体 i 的实际选择，1 代表合作，0 代表背叛，则信誉分变化公式为

$$\Delta c_i = A_i - P_j = A_i - \frac{1}{1 + \mathrm{e}^{-c_j}}$$

5 预测模型

5.1 基本假设

1. 个体进行理性决策，其决策仅参考自己与对手的信誉分。
2. 每个个体面对的收益矩阵都是相同的，即环境以同样标准要求所有人。
3. 除了通过博弈直接获取的期望利益外，个体会关注信誉分的变化，并将其纳入考虑。

5.2 模型建立

现实中，人们会被信誉所左右：社会道德标准要求诚信，高信用的个体更可能在未来获得更多资源。通过信用系统的记录，宏观失信管控也成为可能。对于高分个体，继续增加信誉分并不会使自己的受信任程度显著增加，此时的边际收益是很小的。然而，对于低分个体，由于不被从市场中清除永远是第一目标，再次背叛降低分数的边际成本是很高的。因此，本模型选用反比例型函数（即 $y = \frac{k}{x-x_0}$）描述这一现象。个体 i 由于信誉分变化导致的收益为

$$\pi_{i,1} = \frac{k_i}{c_i - c_{min}} \cdot \Delta c_i$$

其中 $k_i > 0$ 代表个体 i 对信誉分的重视程度。k_i 对于不同个体呈正态分布，因此有平均值 k_m，标准差 k_s。容易发现，k_i 越大则 $|\pi_1|$ 越大，代表对信誉分的变化的重视。当然，总收益还包含本轮博弈直接带来的期望收益 $\pi_{i,2}$，共为 $\pi_i = \pi_{i,1} + \pi_{i,2}$。

考虑个体 i 和 j 的博弈。对 i 而言，合作的期望收益为

$$\begin{aligned}\pi_i^C &= P_j \cdot p + (1-P_j)\cdot q + \frac{k_i}{c_i - c_{min}}(1-P_j) \\ &= \frac{1}{1+\mathrm{e}^{-c_j}}\cdot p + \frac{\mathrm{e}^{-c_j}}{1+\mathrm{e}^{-c_j}}\cdot q + \frac{k_i}{c_i - c_{min}}\cdot\frac{\mathrm{e}^{-c_j}}{1+\mathrm{e}^{-c_j}}\end{aligned}$$

背叛的期望收益为

$$\begin{aligned}\pi_i^D &= P_j \cdot r + (1-P_j)\cdot s + \mathrm{e}^{-\lambda_i \cdot c_i}\cdot K\,(0-P_j) \\ &= \frac{1}{1+\mathrm{e}^{-c_j}}\cdot r + \frac{\mathrm{e}^{-c_j}}{1+\mathrm{e}^{-c_j}}\cdot s - \frac{k_i}{c_i - c_{min}}\cdot\frac{1}{1+\mathrm{e}^{-c_j}}\end{aligned}$$

当 $\pi_i^C > \pi_i^D$ 时个体 i 将选择合作，反之则背叛。记

$$\Delta_i \equiv \pi_i^C - \pi_i^D = \frac{1}{1+\mathrm{e}^{-c_j}}\cdot(p-r) + \frac{\mathrm{e}^{-c_j}}{1+\mathrm{e}^{-c_j}}\cdot(q-s) + \frac{k_i}{c_i - c_{min}}$$

则其正负性代表了个体 i 的决策：正为合作，负为背叛。

5.3 合作持续性

当两个个体 i 与 j，在相遇的第一轮选择合作后（即 $\Delta_i > 0$ 且 $\Delta_j > 0$），合作关系能否持续是一个关键问题。随着博弈的进行，使 Δ 发生改变的因素是 c_i 和 c_j，且它们都是单调递增的（二人的合作导致双方信誉分不断上升）。以个体 i 为

例，考虑 Δ_i 对 c_i 及 c_j 的偏导数

$$\frac{\partial \Delta_i}{\partial c_i} = -\frac{k_i}{(c_i - c_{min})^2}$$
$$\frac{\partial \Delta_i}{\partial c_j} = \frac{\mathrm{e}^{-c_j}}{(1+\mathrm{e}^{-c_j})^2} \cdot (p+s-r-q)$$

不难发现，当 $p+s \leq r+q$ 时，$\mathrm{d}\Delta_i = \frac{\partial \Delta_i}{\partial c_i} \cdot \mathrm{d}c_i + \frac{\partial \Delta_i}{\partial c_j} \cdot \mathrm{d}c_j < 0$ 恒成立，故合作一定以其中一方背叛而结束。不过由于双方无法得知对方的 k 值，无法提前预判在哪一轮时背叛首次发生，故不会发生预先背叛。

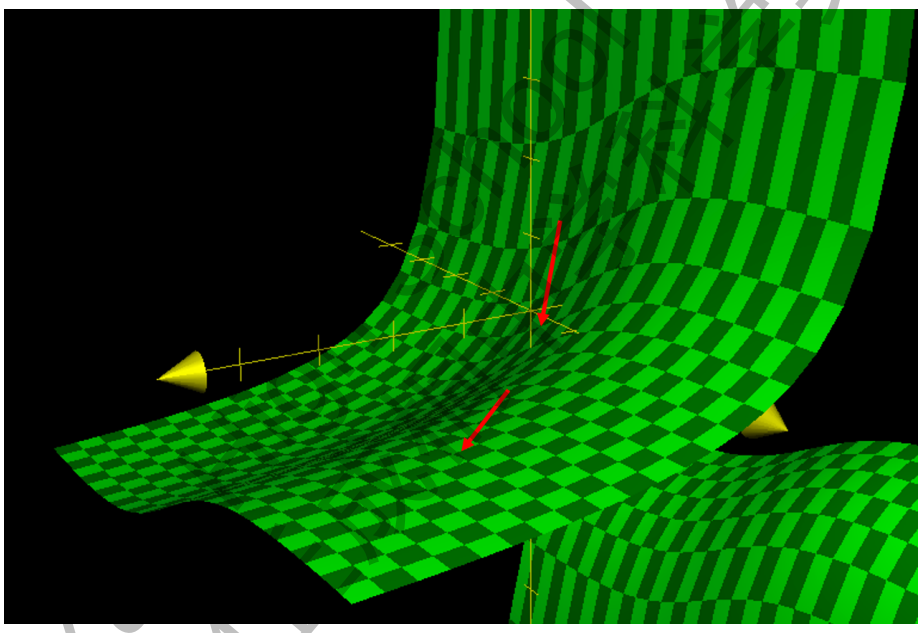

图 1: Δ_i（z 轴）关于 c_i（x 轴）和 c_j（y 轴）的函数图像

如图所示，在 x 轴和 y 轴正方向，Δ_i 递减，并在超越某一阈值后降至 0。此时个体 i 将由合作转为背叛。

具体决策流程如下图，其中绿色代表合作，红色代表背叛。图为 a, b, c, d, e, f

六个个体进行 10 轮博弈的结果。

轮次 \ 个体	a	b	c	d	e	f	合作总人数
1							1
2							5
3							2
4							0
5							6
6							1
7							4
8							2
9							2
10							4

图 2: 个体的决策流程图

6 数据分析与结果

6.1 指标选取

为了有效的证明该信用系统可以促成合作，合作人数，平均收益，平均信誉分作为研究合作程度的指标。

我们发现随着模拟的轮次增加，合作人数，平均收益，平均信誉分数都是收敛的。当 $N = 1000$，$p = 1$，$q = -4$，$r = 3$，$s = -1$，$k_m = 3$，$k_s = 1$，$c_{init} = 0$，$c_{min} = -2$ 时如图所示

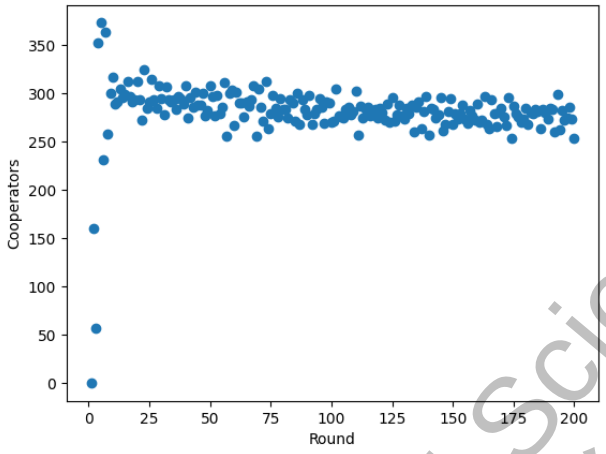

图 3: 合作人数随轮次增加的收敛示意图

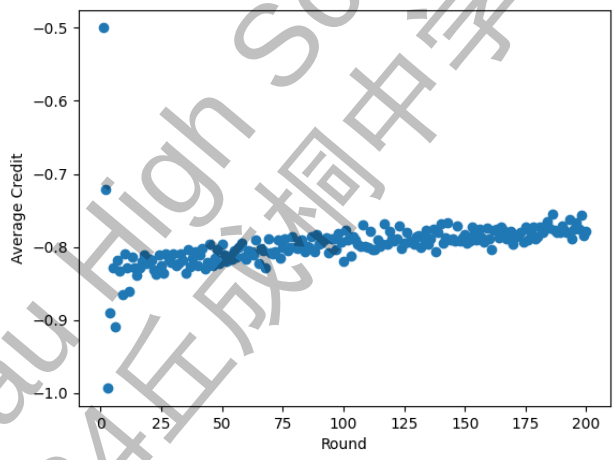

图 4: 平均信誉分随轮次增加的收敛示意图

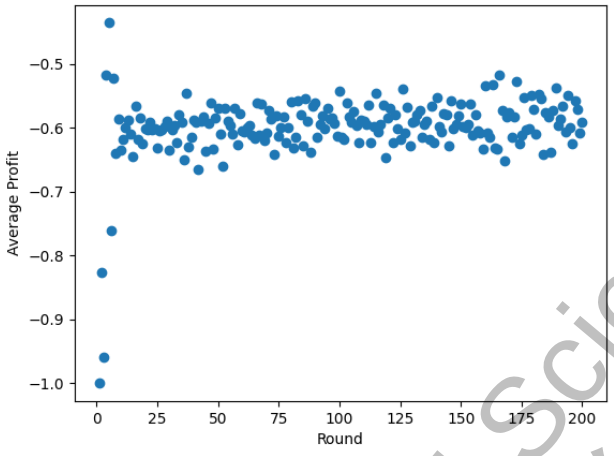

图 5: 平均收益随轮次增加的收敛示意图

发现这三个应变量在一开始时收敛极快，之后有细微变化，但在 200 轮左右区域稳定，因此因此在之后的分析中我们会取这三个量在 190 至 200 轮次的平均值进行分析。

6.2 变量分析

6.2.1 对于随机性的论证

预测模型存在一定的随机性。为了探究随机性带来的影响，令 $p = 1$，$q = -4$，$r = 3$，$s = -1$，$k_s = 1$，$c_{init} = 0$，$c_{min} = -2$ K_m 分别为 3,3.25,3.5 模拟三十次，记录每个指标的最大值，最小值和平均值，有下图。

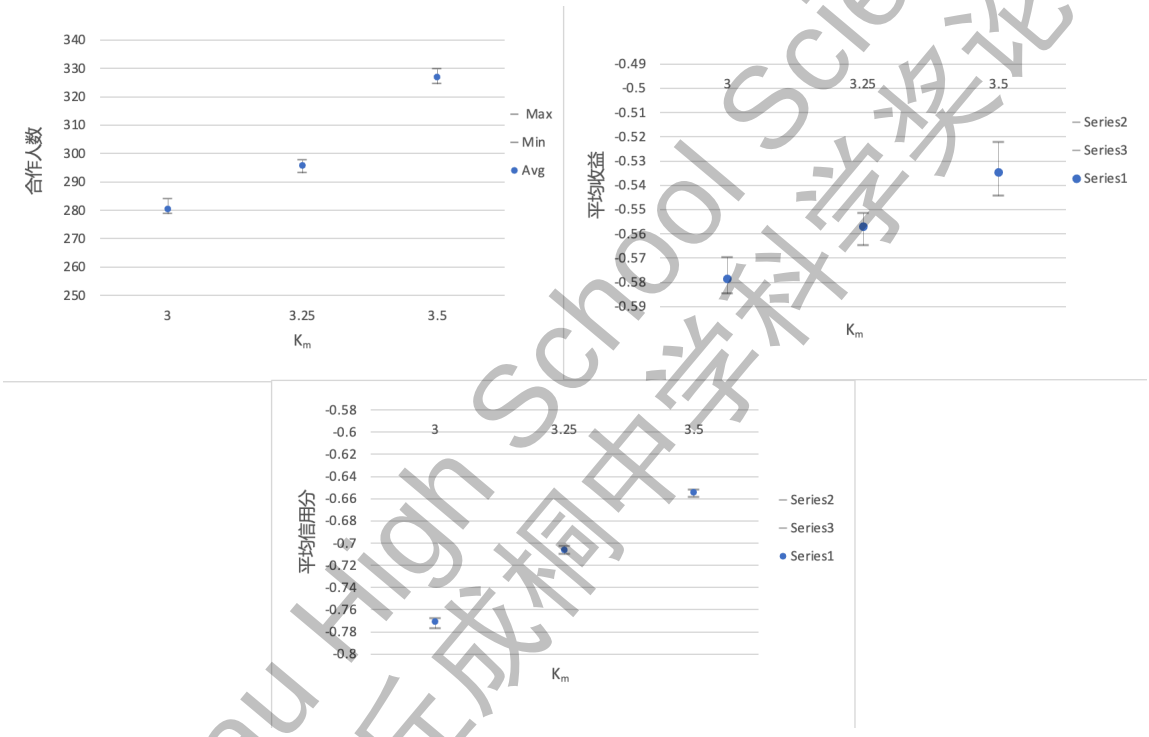

图 6: 平均收益随轮次增加的收敛示意图

可见随机性影响不显著，即使是最为随机的平均收益，其偏差范围仍然不足原有数值的 5%。因此接下来的分析中将会忽略随机性，只进行一次模拟。

6.2.2 对 k_m（信誉分重视程度正态分布的平均值）的分析

当 $p=1$，$q=-4$，$r=3$，$s=-1$，$k_s=1$，$c_{init}=0$，$c_{min}=-2$ 时改变 k_m 得到下列示意图，从左到右，从上到下，分别是 k_m 对合作人数，平均收益，和平均信誉分的影响，最后的图为将三组数据归一化后合并展示。

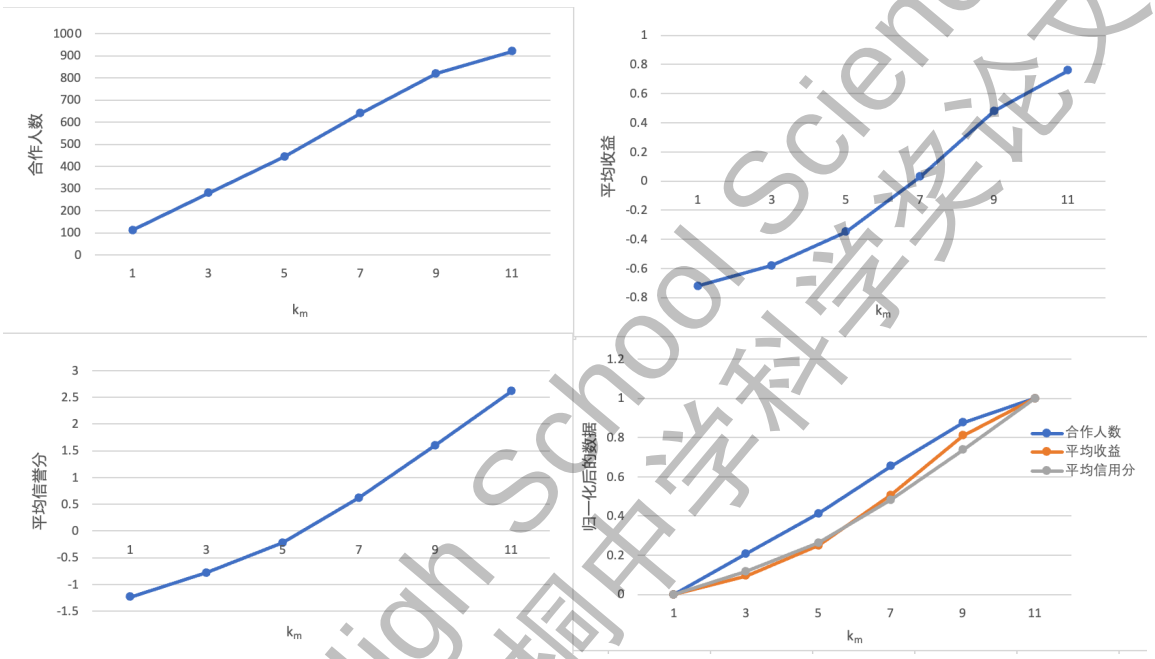

图 7: k_m 变化对于合作人数，平均收益，平均信誉分的影响

可知，k_m 越大，合作就越有可能发生，这也是很符合直觉的，对于信誉分越看重，就越可能合作。

6.2.3 对 k_s（信誉分重视程度正态分布的平均值）的分析

当 $p=1$，$q=-4$，$r=3$，$s=-1$，$k_m=7$，$c_{init}=0$，$c_{min}=-2$ 时改变 k_s 得到下列示意图。

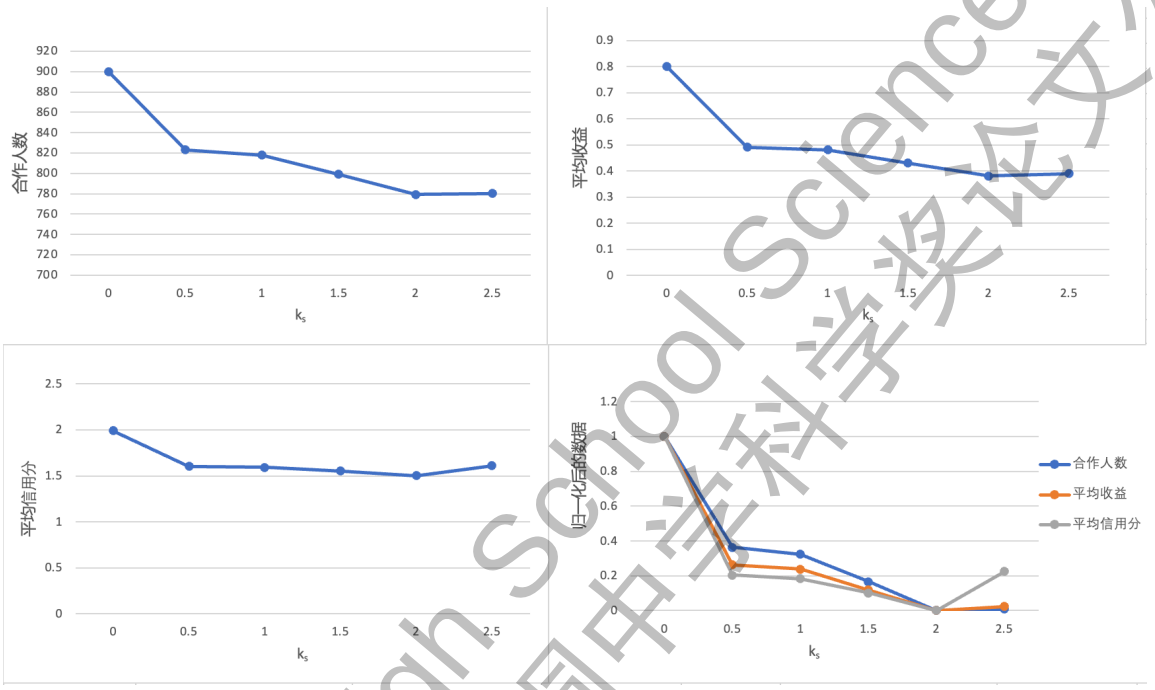

图 8: k_s 变化对于合作人数，平均收益，平均信誉分的影响

可知，随 k_s 越大，合作发生的就越少，说明即使有非常看重信用的人，哪些不看重信用的人也能对整个博弈环境造成负面影响。然而造成的负面影响也是有限的，会逐渐趋近于某一值，并不会使得全员背叛，这是因为信誉分过低的个体会被清除，也反映了清除机制的必要性。

6.2.4 对 c_{init} 的分析

当 $p=1$，$q=-4$，$r=3$，$s=-1$，$k_m=7$，$c_{min}=-2$ 时改变 c_{init} 得到下列示意图

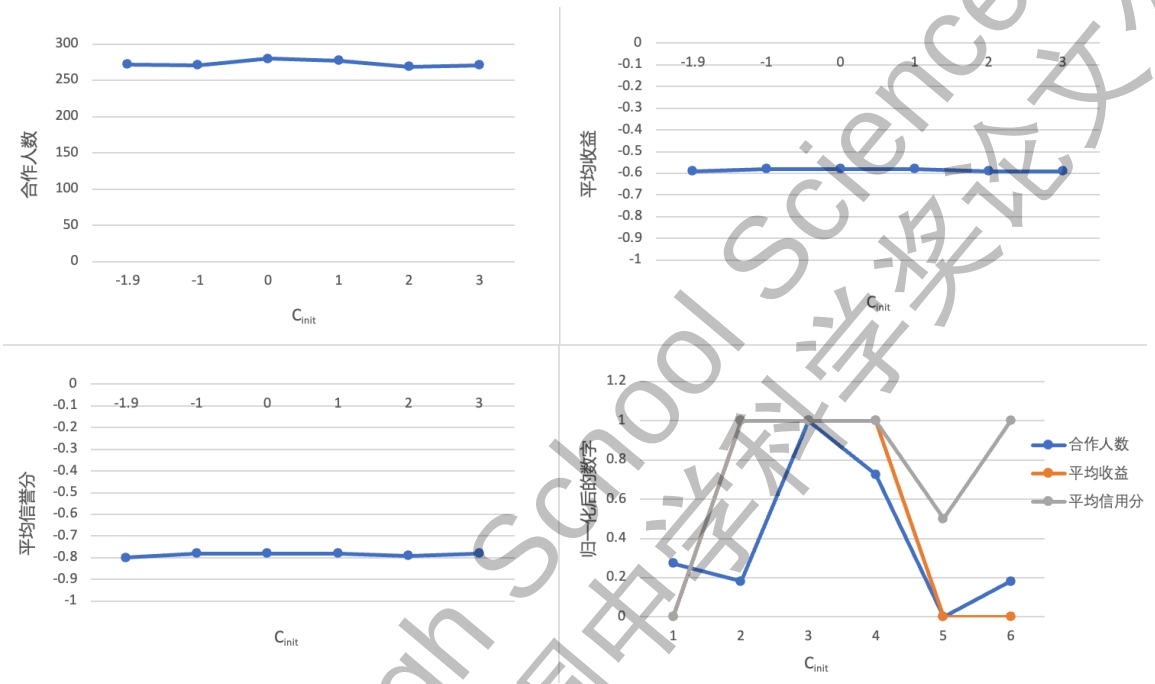

图 9: c_{init} 变化对于合作人数，平均收益，平均信誉分的影响

发现 c_{init} 对合作人数，平均收益，平均信誉影响不显著。

6.2.5 对 c_{min} 的分析

当 $p=1$，$q=-4$，$r=3$，$s=-1$，$k_m=7$，$c_{init}=0$ 时，改变 c_{min} 得到下列示意图

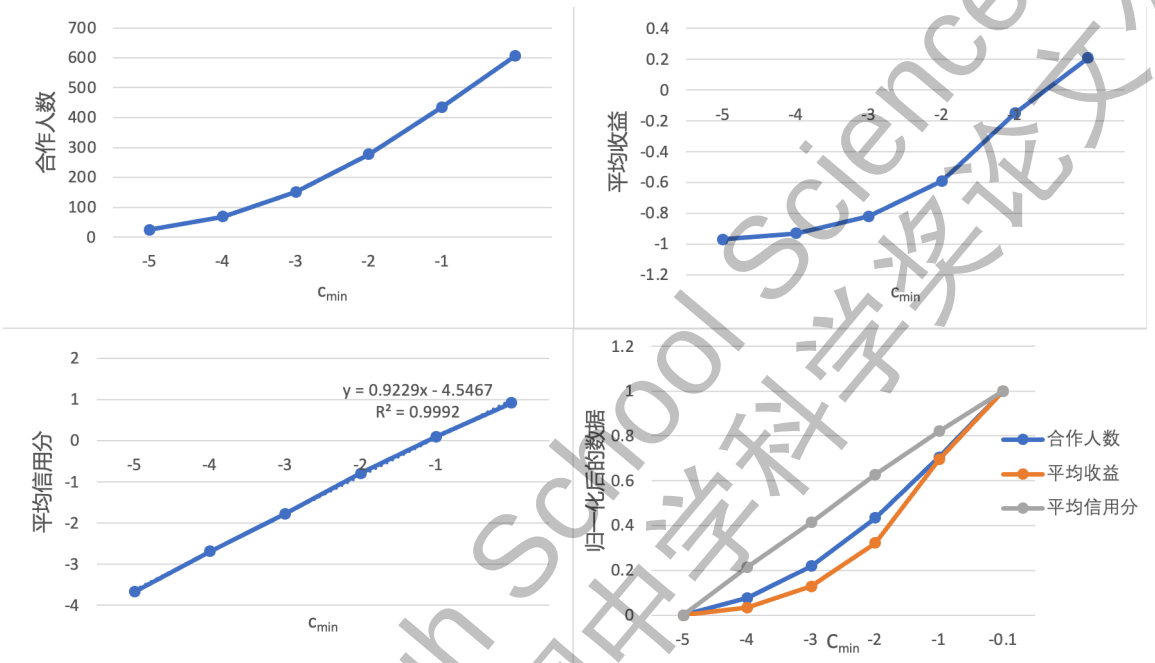

图 10: c_{min} 变化对于合作人数，平均收益，平均信誉分的影响

发现 c_{min} 越高，合作意愿越明显，同时信誉分最后收敛的值和 c_{min} 的取值呈现了相当强的线性关系。用 $y=ax+b$ 拟合后的 r^2 高达 0.9992。

7 结论与启示

通过刚刚分析的四个参数，我们可以得到四条结论。

1. k_m 的取值对于个体之后的合作程度十分重要。可见社会应该多宣传诚实守信之风，以促进合作，增加社会总收益。

2. k_s 增加所带来的随机性会降低整体合作比例，以及总收益，可见不看重信誉分的个体对于整体影响之大，如果有必要社会应该在对于这些个体进行说教或严惩。

3. 参数 c_{init} 对模型表现影响不显著，可以随便设置。

4. 信誉分最后收敛的值和 c_{min} 的取值呈现了相当强的线性关系，可见平均的信誉分会倾向收敛至保证不被清除市场的最低的安全分数。因此可以通过控制 c_{min} 的取来决定整体的合作程度。可以针对不同场景对合作程度的要求来选去不同的 c_{min}，从而在保证合作和维持相对宽松的环境中得到最优解。

8 现实应用及局限

该模型可以应用在部分博弈相关的场景中。例如学校中老师时常布置的小组作业。同组的两个人可以分别选择配合工作或消极应对。此时学生可能陷入囚徒困境，同时选择消极怠工。如果引入类似的信用系统，就可以鼓励同学积极学习，并帮助教师识别消极学生并给予批评教育。

本模型要求信用系统精确地控制每个个体信誉分的增减，但是这通常需要耗费大量的监控、计算资源。在学校班级这种小范围社会内，教师通常可以做到这一点；

但在更大规模的竞争中，这样的监视通常更为困难。这也造成了本模型的主要局限性。

References

Axelrod, R. M. (1984). *The evolution of cooperation*. New York: Basic Books.

Berg, A. (2020). Statistical analysis of the elo rating system in chess. *Chance*, *33*(3), 31–38.

Hadzikadic, M., & Sun, M. (2007). A cas for finding the best strategy for prisoner' s dilemma. In *International conference on computational science*.

Kreps, D. M. (1989). Nash equilibrium. In *Game theory* (pp. 167–177). Springer.

Leinfellner, W. (1986). The prisoner' s dilemma and its evolutionary iteration. In *Paradoxical effects of social behavior: Essays in honor of anatol rapoport* (pp. 135–148). Springer.

Pelánek, R. (2016). Applications of the elo rating system in adaptive educational systems. *Computers & Education*, *98*, 169–179.

Press, W. H., & Dyson, F. J. (2012). Iterated prisoner' s dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, *109*(26), 10409–10413.

Raihani, N. J., & Bshary, R. (2011). Resolving the iterated prisoner' s dilemma: theory and reality. *Journal of Evolutionary Biology*, *24*(8), 1628–1639.

Tucker, A. W. (1950). A two-person dilemma. *Prisoner's Dilemma*.

致谢页

由衷的感谢指导老师王艺臻老师和评委老师们读到这里的耐心。

我们的选题源于经济课上学习的囚徒困境。我们发现在这个情景下，理性的两个人做出选择反而导致了对两个人而言最差的结果。我们就在想为什么会这样，以及如何去避免这种情况。通过查阅网上的资料，我们了解到合作在多轮次的博弈中是有可能产生的，并且了解了各种面对多轮次博弈的“最优策略”。但是，我们没有继续前人的道路为众多“最优策略”再新添一员。与其局限于个体的角度，我们更希望能从整体的角度真正将问题解决，即促使个体选择合作。通过阅读和我们的生活经验，我们觉得促成合作最重要的因素就是信任。如果双方可以快速构建起信任基础，那么囚徒困境就迎刃而解了。于是我们就在想是否可以建立一个足够权威的信用系统，记录下参与者的信用，使得素未谋面的二人可以建立信任，并合作。

开始我们的研究后，我们找到了校内经济老师王艺臻老师作为指导老师，提供无偿指导。之后我们先根据预期收益建立了第一个模型，但是并没有打到我们想要的效果，于是我们和王老师进行了第一次讨论，王老师讲我们可以从马尔可夫链的角度去思考，于是陈昊宇就提出并建立了个体的概率转移矩阵，就是现在文中的引入模型。当齐奕安编好程序后进行模拟，我们还是发现结果差强人意，这时陈昊宇突然想到了之前读过的一本书 *The Evolution of cooperation* 中提到的 Tit-for-tat 策略（即一直模仿对手上一轮的策略），这种方案的优越性在于可以有效的识别并背叛一直倾向背叛的个体，但同时和倾向于合作的个体保持合作。而后我们通过程序模拟发现当大多数人都倾向于 Tit-for-tat 时，不论剩下的人如何长久的合作时可以发生的。因此陈昊宇便得到启示，信用系统应该要鼓励个体进行倾向 Tit-for-tat 的决定。

在这个想法的基础上，齐奕安和陈昊宇共同构建了新的信用系统，并且齐奕安

又根据对人们想法的分析建立了预测模型来测试信用系统。通过程序模拟验证，发现其有不俗的表现。之后陈昊宇完成了信用系统前面的写作，齐奕安完成了信用系统，预测模型构建部分的写作。而后我们又和王老师进行了第二次讨论，王老师给了一些关于论文写作用词，以及论文结构方面的建议。最后，陈昊宇又完成了数据分析可视化和模型结论部分，而齐奕安完成了模型的应用部分，并根据王老师给的建议重新调整了格式。

再次感谢我们的指导老师和评委老师!