自适应博弈视角下的人机混合应急救援团队中的动态协作研究

摘要：?

# 1引言

随着技术的发展，机器人逐渐应用在各个领域，在人类社会中扮演着重要角色。如应用在工业生产的3D分拣机器人[1]，能够抓取不规则的物体并有序放置；医疗领域的康复机器人和手术机器人[2, 3]，能够有效节省医务资源并提高手术精度。同样地，机器人在应对人类灾难时也发挥着重要作用，UAV、UGV、USV等已应用在应急安全和救援场景中。据安全和救援机器人的协会（SSRR）预测，未来十年机器人将在公共安全领域得到广泛应用[4]。机器人能够通过感知外部状态做出合理的行为决策，以帮助人类完成复杂或危险工作。但是只有机器人组成的团队不足以应对突发状况，尤其是在复杂救援场景中。

研究表明，在机器人的团队中加入人的参与，能够有效引导机器人在执行任务过程中的进化行为[5]。因此，人和机器人合作共同完成任务成为一种高效、安全、准确的工作形式。在人和机器人共同组成的团队中，机器人不再是帮助人类完成任务的工具，而是扮演具有感知能力、知识处理能力和自然语言交互能力的队友的角色[6]。人类-机器人的复合团队中的合作已经成为重要的研究方向，亟需应急安全和救援领域的重视。

在人类-机器人的复合团队的合作中，不可避免的面临互相协作（coordination）的问题。即使我们可以简单的认为机器人总具无条件与人合作的意愿，但如果遇到合作意愿较低的人类主体时，协作也很有可能出现问题。和机器人成员不同，团队中人类成员的合作意愿受个体因素和社会因素的影响，不同的合作意愿会不可避免的反映在合作行为上。团队成员共同完成任务时，会出现双方在合作方式上无法达成一致的情况，也会有虽然意愿不同但通过社会学习实现有效协作的可能。当有限理性的人类和完全理性的机器人组成团队时，存在人和人、人和机器人、机器人和机器人三类可能的合作场景，协作关系具有丰富的动态变化。那么对于这样的团队，协作的长期演化过程是怎样的：不同合作意愿的成员能否通过社会学习实现有效协作还是困于持续的协作失败之中？

国内外学者对于应用博弈论进行团队协作的研究已有较长的历史，然而，这些研究并未涉及到机器人和人所构成的复杂混合团队，绝大多数是关注由单一的人类主体构成的团队。而在机器人与人交互领域，研究者们已经意识到了协作问题的存在。Liu & Hind[19]结合交互记忆系统的理论，认为在人机响应时间团队中，人会赋予机器人不同的专业知识，但不会共享交互记忆系统，成为人机团队之间的协调障碍。Murphy[20]提出搜救行动中的协调问题尚未解决，机器人团队的自由度越高，越需要更多专业的操控人员，救援任务的部署更加复杂，所以为降低人和机器人的比例，需要加强团队的协调技术。Iqbal & Riek[21]发现不同的算法会影响人类团队的行为，当两个行为相似的机器人加入人类团队时，团队协调性没有发生变化，加入两个不同算法的机器人时，会降低团队的同步性。然而，他们的工作并没有提供一个探讨机器人与人协作在较长时间尺度上动态变化的分析框架，因而在回答上述问题有其局限性。

针对现有文献的不足，本文借助演化博弈论，将个人的合作意愿表述为博弈模型中的个性化收益，构建了基于协作博弈的自适应博弈过程。通过仿真实验对该过程的长期动态进行了探讨，我们发现，在长期充分博弈的情况下（1）：机器人-机器人构成的合作总是可以实现高效协作，但是机器人在何种策略构成协作惯例具有很大的随机性；（2）：具备不同偏好的人与人构成的合作存在较高的无法实现有效协作的可能，取决于个性化收益与协作收益的数量关系；（3）：在机器人不具备偏好的情况下，机器人-人构成的合作可以实现高效协作，但在机器人使用不同的行为策略选择机制时，协作所带来的团队整体收益可能有所区别。本文的结果突出了个人偏好的重要影响以及对机器人的策略机制进行适应性设计的必要性，为未来的人机混合应急救援团队中的协作研究提供了借鉴。

# 2理论基础

## 2.1协作博弈

协作博弈是对真实世界中协作的抽象。图1描述了最基本的2🞨2静态对称协作博弈。显然，该博弈有两个纯策略纳什均衡，当两个参与者都使用策略*X*或*Y*时，因为其策略一致（*X*，*X*）或（*Y*，*Y*），便在这一策略上形成了协作，两者均可获得一致行动所产生的协作收益1。若协作无法达成，则两者的收益均为0。在本文的情形下，我们假设有两个救援队员，他们有两种策略，一种是共同合作来完成某项任务，一种是各自独立工作，他们实现协作也就意味着他们采用了相同的一致策略：即（合作，合作）或（独立工作，独立工作），这样可以获得协作收益。如果一方选择合作，而另一方选择独立工作（合作，独立工作），或（独立工作，合作），他们就无法实现协作，希望合作的一方无法达成自己的愿望，而希望独立工作的一方也觉得受到另一方的干扰，因此两者的收益都非常低。

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | *Player B* | |
|  |  | *X* | *Y* |
| *Player A* | *X* | *1, 1* | *0, 0* |
| *Y* | *0, 0* | *1, 1* |

图1. 简单的2🞨2静态对称协作博弈。

上述简单模型的一个不足之处在于并未考虑博弈参与者的个人偏好。例如，参与者*A*可能从个人角度出发，偏好行为*X*，无论是否能够与*B*在策略上形成一致（协作成功），使用X均可为其带来某种心理上的满足或其他形式的个性化收益。因此，我们可以对上述简单模型进行扩展，增加A的个性化收益，记做*AX*，*AY*，和B的个性化收益，记做*BX*，*BY*，博弈结构也就相应的变为了图2的结构。如果我们假定个性化收益严格小于协作收益，在一次性博弈的情况下，其纳什均衡仍然是（*X*，*X*）或（*Y*，*Y*）。

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | *Player B* | |
|  |  | *X* | *Y* |
| *Player A* | *X* | *1+AX, 1+BX* | *AX, BY* |
| *Y* | *AY, BX* | *1+AY, 1+BY* |

图2. 包含了个性化收益的2🞨2静态对称协作博弈。

另外，个人偏好的影响不仅仅限于在其策略上构成个性化收益，很多时候也限制了其在某些策略上的选择。让我们考虑三种策略（*X*，*Y*，*Z*）的协作博弈，A的个人偏好限制其只能选择*X*，*Y*两种策略，而B的个人偏好限制其只能选择*Y*，*Z*两种策略。那么，二者可能的策略组合就只有四种：（*X*，*Y*），（*X*，*Z*），（*Y*，*Y*），（*Y*，*Z*）而非九种。其中可能的纳什均衡智能在（*Y*，*Y*）上形成。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Player B* | | | |
|  |  | *X* | *Y* | *Z* |
| *Player A* | *X* | *-* | *AX, BY* | *AX, BY* |
| *Y* | *-* | *1+ AY, 1+BY* | *AY, BZ* |
| *Z* | *-* | *-* | *-* |

图3. 考虑个人策略限制的协作博弈。

为了描述有人和机器人构成的团队，我们需要将模型从简单的两个参与者扩展到多个参与者在一个较长的时间尺度上进行的博弈过程。在这种情况下，简单的采用纳什均衡来对博弈的性质进行描述就有所不足了。在一个多个参与者构成的长期离散时间博弈中，系统的状态可能会非常复杂，此外，由于个人偏好等限制，使得纳什均衡未必可以成为系统的支配性策略，因此，我们更为关注的是，经过长期的充分博弈之后，不同的策略在所有参与者上的分布，同时，我们也关注实现最终策略所经历的动态过程，因此，需要在演化博弈的视角下的新的分析工具，这就是我们将要介绍的自适应博弈。

## 2.2自适应博弈

事实上，个体在决策时受制于其有限理性（可以获取的信息有限，处理信息的能力有限）很难做出最优选择，通常是根据已有的经验决策或通过社会学习模仿他人的行为。Forster & Young[12]、Kandori et al.[13]、Young[7]在研究惯例形成时考虑了参与者有限理性[14]的特点，限制参与者的决策信息和资源，构建了持续随机性的进化博弈模型，用来描述多人组成的博弈达到多重均衡的动态过程。Young[7]将这一过程定义为自适应博弈（Adaptive play）。

自适应博弈基本思想是：每一期随机选取*n*个参与者两两交互，每期参与者通过上期部分参与者的有限行为信息判断当前最优行为，长期多次交互会达到某一严格纯策略均衡，最先达到的均衡取决于过程的不确定性和初始状态。当参与者有新的尝试或犯错时，动态过程会出现不同状态在长期过程中的相对频率的平稳分布，犯错概率很小时，这个平稳分布集中在纯策略纳什均衡特定子集附近。

部分研究尝试将自适应博弈作为研究团队行为的工具。Wu & Wang[15]构建了连续公共物品博弈中重复互动合作演化的模型，发现在折现公共品博弈中，较弱的扰动有利于完全合作状态，较强的扰动对完全合作状态和部分合作状态均有利。Rubin & O’Connor[16]讨论了歧视和学术合作的共同演化，研究表明少数群体更容易受到歧视，歧视会影响不同群体成员与群体内部人员的合作，从而降低合作网络的个人多样性。Hwang[17]将惩罚看作一种协调活动，发现虽然惩罚者之间的协调随小组规模的增加而减弱，但团队规模的优势更大，大规模团队更有利于组织合作。自适应博弈同样适用于现实自主主体的研究，如软件设置和机器人场景中。Stone et al.[18]设计了新旧机器人组成的临时团队共同搜救的场景，试验了同时决策和顺序决策两种方式，均表明掌握了更多环境信息的新机器人会影响旧机器人的行为，以达到团队效用最优。

# 3博弈模型

## 3.1基本假设

3.1.1参与者

设计一个人和机器人组成的救援团队实施救援搜救的场景。团队中有*n*（*n*≥2）名成员，其中人员的数量为*n1*，机器人的数量为*n2*，在救援过程中团队没有成员离开，也没有新的成员加入。团队成员的集合用*N*表示，*i*表示某一具体团队成员。

3.1.2合作意愿

将团队中人类成员的合作意愿划分为*L*个层级{*L1, L2, …, Lj, …, LL*}，每个成员都有所在的层级，*Lj*层级上成员的集合记为*Nj*，∑*j* |*Nj* |= *n1*。如果*Lj*层级上没有成员，则*Nj*=∅。

3.1.3合作行为

团队成员共有*K*种合作行为，每种行为代表一个纯策略，策略集合为*S=*{*s1, s2, …, sk, …, sK*}。团队成员在不同合作意愿下可接受的合作行为不同，合作意愿为*Lj*的成员可接受的策略集合记为*Sj*，*Sj∈S*。设定以下基本假设：

假设1：合作意愿为*Lj*的成员在*t*时刻只能从*Sj*中挑选一种合作行为；

假设2：交互时双方知晓对方的合作意愿层级，但不能确定对方会采取何种合作行为。

同时，若策略集*Sj*∩*Sp* =∅，则表明合作意愿*Lj*和*Lp*不存在形成协作的策略；若|*Sl* |=*K*，表示合作意愿*cwl*的成员能够接纳策略集中的所有行为。

## 3.2效用函数

团队成员的效用函数由个体效用和交互效用两部分组成，结合VNM效用函数理论[22]，团队成员的效用函数表示如下：

*P(si, s-i) = P\_interaction + f(i,s)*

*f(i,k)*是团队成员*i*选择*k*策略的个性化效用函数，表示个体对所在层级中某一策略的独特偏好，和时间以及对方的行为无关。若=*argmax f(i,s)*，那么*si*为成员*i*在其合作意愿下的最优策略。*P\_interaction*是*i*和其他成员的交互效用函数，用来表示双方交互成功与否。按照协作博弈的定义，*si* =*s-i*时表示交互成功，记为1，*si* ≠*s-i*时为交互失败，记为0。

个体对不同策略的偏好差异使得双方合作时存在利益冲突，假设对于所有成员*i*的*f(i,ss)*函数的取值范围是[0, 1)，那么成员之间一次交互可以看做一次协调博弈。这里机器人没有独特的合作偏好，所以对机器人而言*f(i,s)*=*0*。由于关注的是长期的演化过程，因此初期的个体之间在交互中使用的随机试探对长期的协作趋势的影响会逐步被策略性交互所抵消。

## 3.3决策过程

3.3.1决策信息

当团队成员多次进行两两交互时，形成了一个马尔科夫过程，本文在Young[7]的自适应模型的基础上，对这个过程进行描述。

假设时间是离散的，*t=1, 2, 3, …T*。在每个*t*时刻团队，均有两个团队成员被随机挑选出来进行交互，这两个成员可能有相同的合作意愿或是不同的合作意愿。

每个时刻的两个成员在交互中的策略记录（*si, s-i*）在*H(t)*中。对于两个处于合作意愿层级*j*和*p*的个体，其交互可称为*j-p*交互，将*H(t)*中所有属于*j-p*交互的历史记做*hj-p*，显然，*hj-p*是*H(t)*的子集，而*hj-p*的实质就是*j-p*交互中主体的决策依据。

在交互时双方能够知晓彼此的合作意愿，成员通过回顾两个层级之前的交互历史形成先验知识。人是有限理性的，往往根据有限的行为信息做出决策，并不能回溯整个历史。假设时间*t*选中的成员只能处理长度为*m* 的历史，那么*hj-p(m)*就是该成员进行决策的依据，在没有犯错的情况下，该成员利用某种决策模型中选取预期效用最大的行为作为*t*期决策。在本文中，我们假定选定的成员使用最佳回应策略来进行决策。机器人完全理性，能够获取时间*t*内各个层级的所有历史信息*H(t)*，并总是倾向选择和对方一致的合作行为。

3.3.2最佳回应策略

团队成员根据历史信息*hj-p(m)*得到对方决策的概率分布，进而计算得到最佳决策。假设*t*期团队成员*i*和*j*进行交互，双方均有*s1, s2*两种策略可选，*i*根据其有限的行为信息得知*j*选择行为*s1*的概率为*P1*，选择行为*s2*的概率为*P2*，成员*i*计算两种策略的期望效用：*E(s1)=P1U(s1, s1)+P2U(s1, s2)*，*E(s2)=P1U(s2, s1)+P2U(s2, s2)*。若*E(s1)>E(s2)*，成员*i*采取行为*s1*；若*E(s1)<E(s2)*，*i*采取行为*s2*；若*E(s1)=E(s2)*，成员*i*在二者之间随机选择一个合作行为。

**3.4协作惯例**

对于任意的*j-p*交互，如果其历史*hj-p*中绝大多数的交互都实现了有效协作，即*si=s-i=k*，那么无论如何抽样产生的*hj-p(m)*，其主体依然由（*k，k*）构成，这样，可以期望在主体不犯错的情况下，他们将持续的使用策略k，以维持协作的成功，我们将*k*称作*j-p*交互的协作惯例。

**3.2 实例**

假设一个团队由20个成员构成，包括4个机器人（*R1-R4*）和16个人类成员（*P5-P20*）。总共有4种策略可供成员选择（*X*，*Y*，*Z*，*W*）。这16个人类成员的合作意愿可以分为4级，表示为*L1*-*L4*，其中*L1*的合作意愿最强，*L4*的非合作意愿最强。对于某个人类主体*p*，如果其属于*L1*，则他可以使用的策略为（*X*，*Y*）；如果其属于*L2*，则他可以使用的策略为（*X*，*Y*，*Z*）；如果其属于*L3*，则他可以使用的策略为（*Y*，*Z*，*W*）；如果其属于*L4*，则他可以使用的策略为（*Z*，*W*）。下面的矩阵给出了其在这些策略上可能的个性化收益，其中“-”代表该合作意愿层级的成员不会使用的策略。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *X* | *Y* | *Z* | *W* |
| *L1* | *0.8* | *0.4* | *-* | *-* |
| *L2* | *0.4* | *0.8* | *0* | *-* |
| *L3* | *-* | *0* | *0.8* | *0.4* |
| *L4* | *-* | *-* | *0.4* | *0.8* |

3.2.1机器人与机器人之间的交互

机器人不具备个人偏好，因此，任意两个机器人之间的交互（*R-R*）将R-R)形成如下的博弈结构：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *1, 1* | *0* | *0* | *0* |
| *Y* | *0* | *1, 1* | *0* | *0* |
| *Z* | *0* | *0* | *1, 1* | *0* |
|  | *W* | *0* | *0* | *0* | *1, 1* |

3.2.2人与人之间的交互

人与人之间的交互有多种情况，对于任意的两个人类成员，他们那可能具有相同或不同的合作意愿，因而属于不同的层级。在本例中，存在4种层级，不考虑交互主体所属层级的先后顺序，交互的种类也就有4*+C(4,2)=*10种。我们在下面列出了不同的交互所对应的博弈结构。

* *L1*-*L1*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *1.8, 1.8* | *0.8, 0.4* | *-* | *-* |
| *Y* | *0.4, 0.8* | *1.4, 1.4* | *-* | *-* |
| *Z* | *-* | *-* | *-* | *-* |
|  | *W* | *-* | *-* | *-* | *-* |

* *L1*-*L2*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *1.8, 1.4* | *0.8, 0.8* | *0.8, 0* | *-* |
| *Y* | *0.4, 0.4* | *1.4, 1.8* | *0.4, 0* | *-* |
| *Z* | *-* | *-* | *-* | *-* |
|  | *W* | *-* | *-* | *-* | *-* |

* *L1-L3*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *-* | *0.8, 0* | *0.8, 0.8* | *0.8, 0.4* |
| *Y* | *-* | *1.4, 1* | *0.4, 0.8* | *0.4, 0.4* |
| *Z* | *-* | *-* | *-* | *-* |
|  | *W* | *-* | *-* | *-* | *-* |

* *L1-L4*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *-* | *-* | *0.8, 0.4* | *0.8, 0.4* |
| *Y* | *-* | *-* | *0.4, 0.4* | *0.4, 0.8* |
| *Z* | *-* | *-* | *-* | *-* |
|  | *W* | *-* | *-* | *-* | *-* |

* *L2-L2*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *1.4, 1.4* | *0.4, 0.8* | *0.4, 0* | *-* |
| *Y* | *0.8, 0.4* | *1.8, 1.8* | *0.8, 0* | *-* |
| *Z* | *0, 0.4* | *0, 0.8* | *1, 1* | *-* |
|  | *W* | *-* | *-* | *-* | *-* |

* *L2-L3*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *-* | *0.4, 0* | *0.4, 0.8* | *0.4, 0.4* |
| *Y* | *-* | *1.8, 1* | *0.8, 0.8* | *0.8, 0.4* |
| *Z* | *-* | *0, 0* | *1, 1.8* | *0, 0.4* |
|  | *W* | *-* | *-* | *-* | *-* |

* *L2-L4*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *-* | *-* | *0.4, 0.4* | *0.4, 0.8* |
| *Y* | *-* | *-* | *0.8, 0.4* | *0.8, 0.8* |
| *Z* | *-* | *-* | *1, 1.4* | *0, 0.8* |
|  | *W* | *-* | *-* | *-* | *-* |

* *L3-L3*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *-* | *-* | *-* | *-* |
| *Y* | *-* | *1, 1* | *0, 0.8* | *0, 0.4* |
| *Z* | *-* | *0.8, 0* | *1.8, 1.8* | *0.8, 0.4* |
|  | *W* | *-* | *0.4, 0* | *0.4, 0.8* | *1.4, 1.4* |

* *L3-L4*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *-* | *-* | *-* | *-* |
| *Y* | *-* | *-* | *0, 0.4* | *0, 0.8* |
| *Z* | *-* | *-* | *1.8, 1.4* | *0.8, 0.8* |
|  | *W* | *-* | *-* | *0.4, 0.4* | *1.4, 1.8* |

* *L4-L4*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *-* | *-* | *-* | *-* |
| *Y* | *-* | *-* | *-* | *-* |
| *Z* | *-* | *-* | *1.4, 1.4* | *0.4, 0.8* |
|  | *W* | *-* | *-* | *0.8, 0.4* | *1.8, 1.8* |

上述10种交互可以分为三类。第一类为同阶层交互，包含*L1*-*L1*交互，*L2-L2*交互，*L3-L3*交互，*L4-L4*交互；第二类为阶层不同，但存在建立协作惯例可能性的交互，包括*L1*-*L2*交互，*L1*-*L3*交互，*L2*-*L3*交互，*L2*-*L4*交互，*L3*-*L4*交互；第三类为不存在建立协作惯例可能性的交互，这类只有一种，为*L1*-*L4*交互。

3.2.3机器人与人的交互

显而易见，如果在机器人与人的交互中，由于机器人没有偏好，所有可以使用所有的交互策略，其博弈结构可以简单的分为四种，分别对应四个合作意愿层级。

* *R-L1*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *1, 1.8* | *0, 0.4* | *-* | *-* |
| *Y* | *0, 0.8* | *1, 1.4* | *-* | *-* |
| *Z* | *0, 0.8* | *0, 0.4* | *-* | *-* |
|  | *W* | *0, 0.8* | *0, 0.4* | *-* | *-* |

* *R-L2*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *1, 1.4* | *0, 0.8* | *0, 0* | *-* |
| *Y* | *0, 0.4* | *1, 1.8* | *0, 0* | *-* |
| *Z* | *0, 0.4* | *0, 0.8* | *1, 1* | *-* |
|  | *W* | *0, 0.4* | *0, 0.8* | *0, 0* | *-* |

* *R-L3*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *-* | *0, 0* | *0, 0.8* | *0, 0.4* |
| *Y* | *-* | *1, 1* | *0, 0.8* | *0, 0.4* |
| *Z* | *-* | *0, 0* | *1, 1.8* | *0, 0.4* |
|  | *W* | *-* | *0, 0* | *0, 0.8* | *1, 1.4* |

* *R-L4*交互

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Player B* | | | | |
|  |  | *X* | *Y* | *Z* | *W* |
| *Player A* | *X* | *-* | *-* | *0, 0.4* | *0, 0.8* |
| *Y* | *-* | *-* | *0, 0.4* | *0, 0.8* |
| *Z* | *-* | *-* | *1, 1.4* | *0, 0.8* |
|  | *W* | *-* | *-* | *0, 0.4* | *1, 1.8* |

# 4仿真设计

利用3.2中的数值实例，我们通过仿真对该博弈过程的动态性质进行讨论。在仿真实验中所使用的主要参数如下表所示：

|  |  |
| --- | --- |
| **参数** | **仿真取值** |
| 团队人数 | 20 |
| 机器人主体 | 4 |
| 人类主体 | 16 |
| 博弈策略 | 4种策略：*X*，*Y*，*Z*，*W*；*X*：充分合作，*W*：完全不合作 |
| 人类主体合作意愿 | 4个层级：*L1*-*L4*，其中*L1*的合作意愿最强，*L4*的非合作意愿最强 |
| 合作意愿对主体策略的限制 | 见3.2 |
| 博弈收益结构 | 由协作收益和个性化收益两部分构成，见3.2 |
| 总时间 | 2000次交互 |
| 人类主体记忆大小 | 3 |
| 任一种交互起始随机试探次数 | 5次 |
| 扰动率 | 5% |

在2000个时间单元中，任一时刻*t*，都有两个团队成员被随机挑选出来进行交互。在上文的数值实例中可以发现，不同种类的交互之间是相互独立的，因此，在仿真实验中，在某个特定的时刻*t*，只需关注在在一时刻进行交互的两者所属的层级，就可以将其归类到对应种类中去。为了更好地描述动态系统的性质，在下一部分中，结果将以各独立种类的方式分别进行报告。

**5 仿真实验结果**

在本部分中，我们对仿真实验的结果进行了总结。我们将按照交互主体的不同，按照（1）机器人-机器人交互，（2）人-人交互，（3）机器人-人交互的顺序依次报告仿真实验结果。并对仿真实验中发现的规律做了进一步探讨。

**5.1 机器人-机器人之间的交互**

从理论角度来看，机器人与机器人之间的交互实质上等同于两个完全同质的参与者进行的博弈，同时，由于假设中机器人对过去交互历史进行记忆和处理的能力不受限制，因此，机器人-机器人之间一定可以实现有效的协做。同时，机器人-机器人之间达到有效协作的演化速度非常快，只要早期的随机试探过程中出现了协作成功的交互，那么协作惯例就可以自然的建立起来。仿真结果显示实现所有机器人-机器人之间协作惯例一般在我们预设的随机试探阶段后可立即实现（图）。然而，考虑到机器人所占的比例较小，总体的协作惯例可能依然需要在整个系统上数百次交互（包含所有交互种类）之后才能建立，因此，一个较好的策略是提前让机器人进行交互，预先达成协作惯例。此外，由于在何种策略上达成协作惯例取决于早期的随机试探阶段，因此有必要从设计应急救援机器人之初就进行干预。

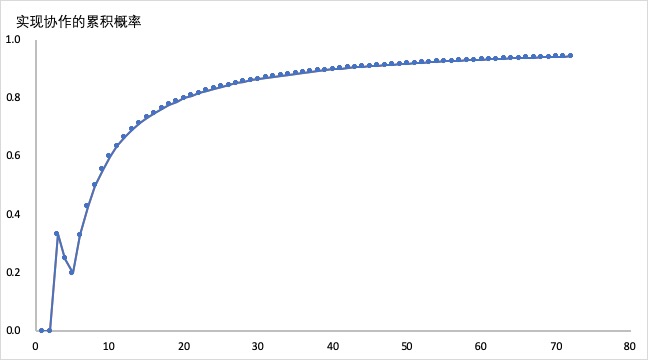


图-. 机器人-机器人之间交互实现协作的累积概率。

**5.2 人-人之间的交互**

人与人之间的交互共有10种情况，我们按照三类不同的交互来对实验结果进行讨论。

5.2.1 同一层级之间的交互（*L1-L1*，*L2-L2*，*L3-L3*，*L4-L4*）

和机器人与机器人之间的交互类似，同一层级之间的交互，可以被认为是两个完全同质的人类主体进行的交互，因为其策略选择和个性化收益均一致，所以很容易形成协作惯例。仿真实验的结果也证实了这一点（图）。我们可以看到，即使在存在随机扰动的情况下，也只是对累积概率稍有影响，总体的协作可迅速回归惯例状态。

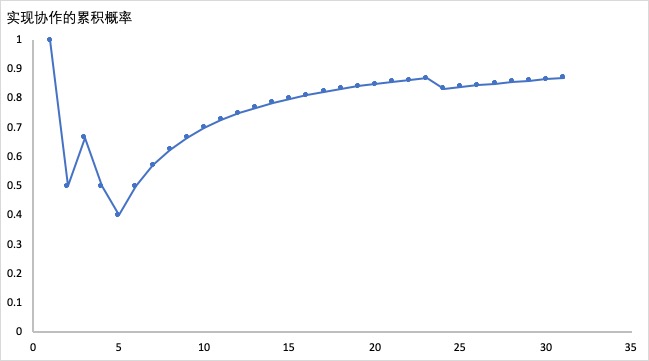


图-. *L4-L4*之间交互实现协作的累积概率。

5.2.2 无构成协作惯例可能的不同层级之间的交互（*L1-L4*）

对于这种情形，显然将不得不陷入持续性的协作失败的境地。*L1*与*L4*之间的交互就是如此，在*L1-L4*交互中，属于L1的主体只能使用X，Y这两种策略，而属于L4的主体只能使用Z，W这两种策略，并无交集。这也是所有情况中唯一的一种完全没有纯策略纳什均衡的交互组合，因而不具备形成协作惯例的可能。由于一个典型的*L1-L4*交互过程中实现协作的累积概率永远为0，表现为与X-坐标轴重合的一条直线，本文就不在此绘制其变化图了。

5.2.3 有构成协作惯例可能的不同层级之间的交互

A．*L1-L2，L3-L4*

我们首先来看*L1-L2，L3-L4*这两种交互，这两种交互的共同特点是交互双方某一方（*L1* 或*L4*）的可选策略是（*L2* 或*L3*）的可选策略的真子集，这样，协作惯例就有可能在*L1* 或*L4*的任意策略上形成。直觉上，这两类交互应该有很大可能会迅速达成协作惯例，然而事实可能并非如此。和前面类似，图- 以实现协作累积概率的形式给出了*L1-L2*交互的仿真实验结果。

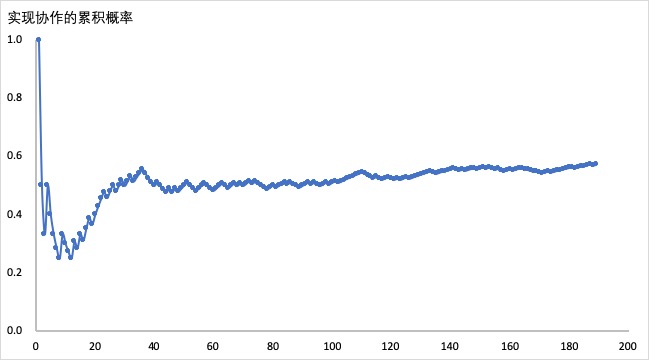
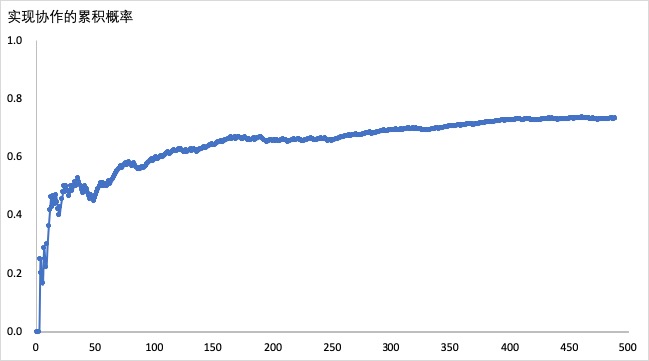


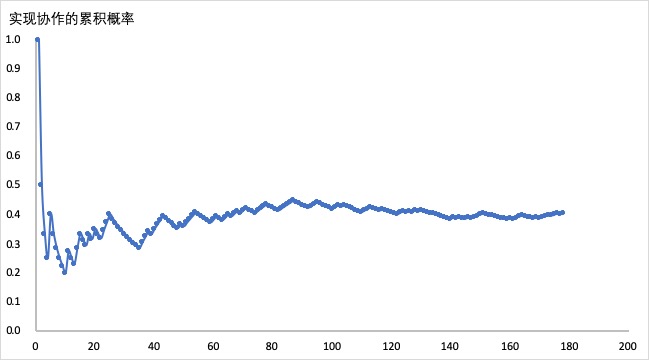
图-. *L1-L2*之间交互实现协作的累积概率。

我们可以看出，在经过接近200次交互后，*L1-L2*交互仍未达到趋近1的累积概率。然而，其累积概率呈现出缓慢增长的趋势。因此，有可能会最终达成协作惯例。为此，我们将仿真的总时间由2000扩展到5000。图- 描述了累积概率的变化，显然，这种缓慢的增长还在持续。进一步的仿真发现，*L1-L2*交互将在约1200次交互后建立协作惯例，无疑，这个时间长度很有可能超出团队合作的整个周期，即使能够达到，意义也相对不大。



B．*L2-L3*

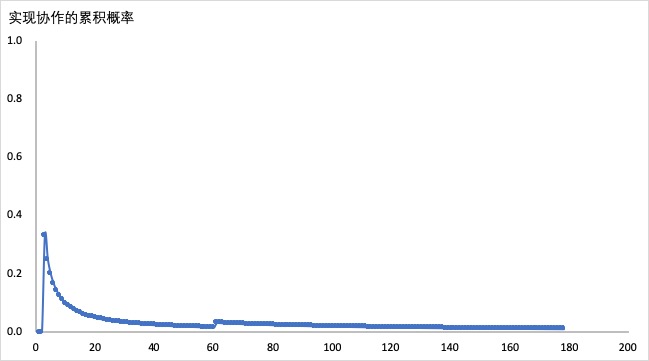
显然，在*L2-L3*的情况中，和前面一样，简单的纳什均衡不足以描述系统的动态性质。我们可以看到，在*L2-L3*交互中，存在两种纯策略纳什均衡，分别是（*Y*，*Y*）和（*Z*，*Z*），在理想状况下，这种交互可以在这两种策略上达到某种协作惯例，然而，仿真的结果却指出，从长远来看，尽管不像*L1-L4*交互那种陷入持续协作失败的境地，也大概只有不到40%的交互会协作成功，远远小于无法协作成功的概率。和A的不同之处在于，*L2-L3*交互中，我们没有看到协作成功的累积概率缓慢增长的趋势，说明*L2-L3*交互可能很难达成协作惯例，而最终在不稳定的状态上反复波动。



图：*L2-L3*交互中实现协作的累积概率变化趋势。

C. *L1-L3, L2-L4*

我们来看最后一种情况，在这种情况里，参与交互的双方只在一种策略上形成协作惯例的可能，同时，协作成功所带来的的收益相对于每个个体的个性化收益的改善并不显著，例如，在中，对属于的个体而言，协作成功的收益改善仅有0.2，可能并不足以驱动其走出舒适区。因此，*L1-L3*和*L2-L4*这两种交互，有可能无法达成协作惯例，甚至会陷入持续性的协作失败。仿真实验的结果也证实了这一点（图-）。在初始随机试探阶段还能偶然的实现协作之后，*L1-L3*交互便迅速陷入了持续性的协作失败中。



图：*L1-L3*交互中实现协作的累积概率变化趋势。

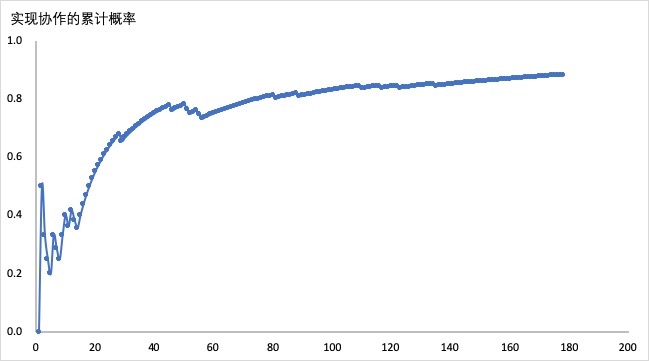
**5.3 机器人-人之间的交互**

在机器人和人的交互中，机器人不受有限理性的限制，能够处理所有的交互历史，其行为也不受个体意愿的影响。同时，处于所有合作意愿层级的人类主体所可以使用的策略集均属于机器人策略集的子集。因此，从理论上讲，机器人总是具备调整自己策略以在长远角度形成协作惯例的能力，从而实现有效协作。仿真实验的结果也证实了这一点。为了保持文章的简洁，我们就不再绘制累积概率变化趋势图了。

在交互的早期，机器人除进行随机试探外，可以选择其他不同的决策方式。常见的有两种，一种为最优社会性决策，即机器人总是使用代表合作意愿最为强烈的策略*X*；另一种为适应性决策，即机器人主动迎合人类主体的偏好，使用对应主体获得最大个性化收益的策略。显然，从团队收益的角度看，后者要比前者更好。在前者中，机器人将很难与*L3*，*L4*层级上的主体实现协作，与*L2*的协作也会在被局限在收益较低的策略上形成协作惯例。我们在机器人与L2人类主体上的交互情况进行了模拟，发现使用适应性决策的总体收益要比使用最优社会性决策的总体收益高出近30%。因此，设计机器人时，采用适应性决策作为交互的起始点是较好的选择。

**5.4 进一步的讨论**

在5.2.3中，仿真实验结果指出，即使存在纳什均衡，协作惯例也可能无法达成，*L2-L3*交互便是这样的例子。根据Basin of Attraction理论，我们假设这类情况是由于交互主体的个性化收益较大，因此往往困于持续使用最大化其个性化收益的策略造成的。为了检验是否如此，我们对个性化收益进行改变，将（0.8，0.4，0）的个性化收益改为较小的（0.2，0.1，0）。这样，如果*L2-L3*交互能够达到协作惯例，较大的个性化收益不利于协作惯例的达成。



图：*L2-L3*交互在较低个性化收益下实现协作的累积概率变化趋势。

利用新的收益结构，我们重新进行了一组仿真实验，结果如图-所示。*L2-L3*交互将能够达成协作惯例，而且时间成本也相对较低。这一点证实了我们的假设。在现实情境中，这意味着团队应该尝试某些机制降低团队成员的个性化收益，通过团队建设放大团队目标相对个人偏好的影响或许是可行的方法。

**6 结论与启示**

本文面向机器人和人构成的复杂团队中的协作问题，利用演化博弈中的自适应博弈方法，考虑到个人主体的不同合作意愿，对协作的长期趋势进行了分析。结果表明：

1. 在机器人与机器人的协作中，协作惯例可以在较短的时间内形成，从而保证了双方远期收益的稳定。然而在何种策略上形成协作惯例具有不确定性。依赖于在交互早期的试探。
2. 在人与人的协作中，协作的长期趋势呈现出复杂的动态状况。当参与协作的双方来自同一合作意愿层级时，协作惯例可以在较短的时间内形成。然而，当参与协作的双方来自不同合作意愿层级时，团队的长期协作趋势可能呈现出多种状态，即使参与协作的双方在特定策略（纯策略纳什均衡）上存在形成协作惯例的可能，这种协作惯例的形成可能要花费巨大的时间成本（*L1-L2，L3-L4*）也未必可以达成（*L2-L3*），甚至有可能陷入持续协作失败的困境（*L1-L3, L2-L4*）。
3. 在机器人与人的协作中，协作惯例可以在较短的时间内形成，然而，通过改变机器人的策略选择模式，可以导致不同的协作惯例，进而导致不同的团队总体收益。

基于本文研究的结论得到如下启示：第一，在具有不同个体特点的复杂人机混合团队中，纳什均衡往往不能充分的描述和解释团队整体协作的长期变化趋势，因此有必要借助类似自适应博弈这类技术对团队的动态演化趋势进行探讨。第二，即便是在人机混合团队中，人与人之间的交互，依然是形成有效协作的主要瓶颈。而个体较强的偏好，往往是形成有效协作惯例的障碍。因此，应该鼓励团队成员形成共同的目标以降低个人偏好的影响。第三，机器人的设计者应该在设计机器人的决策机制时，对决策的长期影响进行考量，并提高机器人在面对不同人类主体时的适应性。

本文研究在一定意义上揭示了机器人和人构成的复杂混合团队中协作的长期演化趋势，为这类团队中的机制设计和机器人的设计提供了一定参考依据。但是本文仍存在不足: 一方面，在建模团队和团队主体的交互时，并未将团队成员构成的网络的拓扑结构考虑在内，只将其划分成机器人与人两类有差别群体，具有一定局限性; 另一方面，个人存在主动改变自己协作意愿的可能，在本文没有进行全面的探讨。今后研究可以从以下几个方面展开（1） 将博弈主体置于不同的网络结构中，进一步探讨人机混合救援团队中的长期协作趋势；（2）充分考虑博弈主体主动改变自己协作意愿的可能，研究如何推动其协作意愿向提升团队收益的方向转变；（3）成员不断变化的开放临时人机混合救援团队中的协作研究等。

# 参考文献

[1] Lin Y, Zhou H, Chen M, et al. Automatic sorting system for industrial robot with 3D visual perception and natural language interaction[J]. Measurement and Control, 2019, 52(1-2): 100-115.

[2] Gomes P. Surgical robotics: Reviewing the past, analysing the present, imagining the future[J]. Robotics and Computer-Integrated Manufacturing, 2011, 27(2): 261-266.

[3] Krebs H I, Volpe B T. Rehabilitation Robotics[M]. // BARNES M P, GOOD D C. Handbook of Clinical Neurology. Elsevier, 2013: 283-294.

[4] Murphy R R, Kleiner A. A community-driven roadmap for the adoption of Safety Security and Rescue Robots[C]//2013 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), 21-26 Oct. 2013. 1-5.

[5] Delmerico J, Mintchev S, Giusti A, et al. The current state and future outlook of rescue robotics[J]. Journal of Field Robotics, 2019, 36(7): 1171-1191.

[6] Seeber I, Bittner E, Briggs R O, et al. Machines as teammates: A research agenda on AI in team collaboration[J]. Information & Management, 2020, 57(2): 103174.

[7] Young H P. The evolution of conventions[J]. Econometrica: Journal of the Econometric Society, 1993, 61(1): 57-84.

[8] Young H P. The Economics of Convention[J]. Journal of Economic Perspectives, 1996, 10(2): 105-122.

[9] Matsui A, Matsuyama K. An Approach to Equilibrium Selection[J]. Journal of Economic Theory, 1995, 65(2): 415-434.

[10] Zhang P, Kang X, Liu Y, et al. Cooperative Willingness Aware Collaborative Caching Mechanism Towards Cellular D2D Communication[J]. IEEE Access, 2018, 6: 67046-67056.

[11] Li B, Ding J, Wang J, et al. Key factors affecting the adoption willingness, behavior, and willingness-behavior consistency of farmers regarding photovoltaic agriculture in China[J]. Energy Policy, 2021, 149: 112101.

[12] Foster D, Young P. Stochastic evolutionary game dynamics[J]. Theoretical Population Biology, 1990, 38(2): 219-232.

[13] Kandori M, Mailath G J, Rob R. Learning, Mutation, and Long Run Equilibria in Games[J]. Econometrica, 1993, 61(1): 29-56.

[14] Selten R. Features of experimentally observed bounded rationality[J]. European Economic Review, 1998, 42(3): 413-436.

[15] Wu T, Wang L. Adaptive play stabilizes cooperation in continuous public goods games[J]. Physica A-statistical Mechanics and Its Applications, 2018, 495(1): 427-435.

[16] Rubin H, O’connor C. Discrimination and Collaboration in Science[J]. Philosophy of Science, 2018, 85(3): 380-402.

[17] Hwang S-H. Conflict technology in cooperation: The group size paradox revisited[J]. Journal of Public Economic Theory, 2017, 19(4): 875-898.

[18] Stone P, Kaminka G, Kraus S, et al. Teaching and leading an ad hoc teammate: Collaboration without pre-coordination[J]. Artificial Intelligence, 2013, 203: 35-65.

[19] Liu L, Hinds P J. Transactive Memory Systems: A perspective on coordination in human-robot incident response teams[C]//2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 11-13 March 2009. 269-270.

[20] Murphy R R, Tadokoro S, Kleiner A. Disaster Robotics[M]. // SICILIANO B, KHATIB O. Springer Handbook of Robotics. Cham: Springer International Publishing, 2016: 1577-1604.

[21] Iqbal T, Riek L D. Human coordination dynamics with heterogeneous robots in a team[C]//2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 7-10 March 2016. 619-620.

[22] Von Neumann J, Morgenstern O. Theory of Games and Economic Behavior[M]. Princeton: Princeton University Press, 1996.