

确定型多局域世界图中 Agent 最优策略解析^①

郑 湛¹, 郑小京^{2*}, 徐绪松³

(1. 武汉纺织大学, 武汉 430073; 2. 哈尔滨商业大学, 哈尔滨 150076; 3. 武汉大学, 武汉 430072)

摘要: 在短的 time-scale 中, 在一个确定的系统拓扑结构下, 用解析的方法确定复杂管理系统 Agent 的最优策略. 构建了一个确定型多局域世界图博弈模型, 对确定型多局域世界图, 分两种情况——局域内 Agent 的合作博弈和局域之间的非合作博弈, 讨论 Agent 行为均衡解; 并将局域内 Agent 的合作随机微分博弈与局域 Super-Agent 之间的非合作随机微分博弈进行了耦合, 以确定复杂管理系统 Agent 的最优策略. 最后讨论了最优策略的稳定性.

关键词: 确定型; 多局域世界图; Agent 的最优策略; 随机微分博弈

中图分类号: F273.6 **文献标识码:** A **文章编号:** 1007-9807(2016)01-0035-12

0 引 言

21 世纪, 复杂的管理系统普遍存在于现实生活中, 如房地产、供应链、区域经济、金融市场等, 这类复杂管理系统, 具有如下特征: 1) 系统具有一定的拓扑结构, 人的行为与系统拓扑结构共同演化. 2) 系统由若干个相互作用的子系统构成, 每一个子系统的主体都为了一个共同的目标而进行合作, 而不同的子系统之间又为了自身利益最大化而竞争. 3) 当时间短时(短的 time-scale), 系统中各个主体间的相互作用发生在一个确定的系统拓扑结构中, 在系统中存在着合作与竞争两种情况, 经过相互作用, 系统中的成员都能找到一个最优策略. 4) 随着时间的增长, 系统自身的特性在不断变化, 为了提升自身的综合竞争能力以及盈利能力, 或与其他主体创建新的关系, 或断开现有的联系. 当时间长时(长的 time-scale), 主体的最优策略将随着一个确定拓扑结构到下一个确定拓扑结构不断发生变化, 没有一个确定的最优策略, 但是, 最优策略的分布特征可能存在.

鉴于复杂管理系统的复杂性, 学者主要采用

仿真的方法对复杂管理系统中的最优策略进行探索, 分析系统特征以及系统与环境之间的相互关系, 并获得了一定的结论^[1-5]. 由于仿真对于系统内部运行机制的处理, 以及规则的发现具有一定的局限性, 为了解决这一问题, 笔者将基于人的行为的复杂管理系统抽象成多局域世界图(称为 Super-Agent, 其中的决策主体为 Agent), 采用随机微分博弈, 研究系统中 Agent 的最优策略^[6-8]. 本文主要研究具有短的 time-scale 特征的、确定的系统拓扑结构下的 Agent 最优策略. 长的 time-scale 特征的、变化系统结构下的 Agent 最优策略的分布特征将在下篇论文中研究.

本文的研究思路是: 首先给出预备知识, 然后对确定的系统拓扑结构构建一个确定型多局域世界图博弈模型, 再对确定型多局域世界图, 分两种情况——局域内 Agent 的合作博弈和局域之间的非合作博弈, 讨论 Agent 行为均衡解, 获得合作博弈的最优策略, 非合作博弈的 Nash 均衡策略. 并将局域内 Agent 的合作随机微分博弈与局域 Super-Agent 之间的非合作随机微分博弈进行耦合, 即将各个局域收益进行合理分配. 最后讨论了

① 收稿日期: 2014-08-12; 修订日期: 2015-10-28.

基金项目: 国家自然科学基金资助项目(71503188; 71040001; 70771083).

通讯作者: 郑小京(1975—), 男, 陕西白水人, 博士, 副教授. Email: tkwiloi_75@126.com

最优策略的稳定性.

1 预备知识

1.1 定义

图的子图 $C_i \in (C_1, C_2, \dots, C_m)$ 被定义为局域(C_i 实际上是一个随机矩阵)^[9]. 如果

$$1) \theta(C_i) = o(N_i \lg N_i) \ll O(N^2) \text{ 其中 } \theta(C)$$

为图 C 的秩, $\sum_{i=1}^m N_i = N$, 其中 N_i 为局域 i 内的 Agent 数量, N 为复杂管理系统中 Agent 的数量.

$$2) \text{ 对于任意 } 1 \leq i, j \leq s, i \neq j,$$

$$C_i \cap C_j = \emptyset, \cup_i C_i = C.$$

1.2 假设

假设 1 假设基于人的行为的复杂管理系统由若干个子系统构成, 这些子系统称之为局域. 这些局域足够大, 使得每一个局域中可以容纳足够数量的决策主体; 同时, 假定这些局域又足够小, 使得有足够数量的局域在不断进行交互作用.

假设 2 在一个短的时间范围(time-scale) 内各个局域中的 Agent 之间以合作博弈机制进行相互作用, 在追求区域整体利益最大化的基础上将收益进行合理分配; 而对于不同的、相互联系的局域, 他们之间按照非合作博弈机制进行相互作用.

对应于确定的拓扑结构的 Agent 合作/非合作博弈, 系统的 Agent 行为有一个动力学方程, 该方程涉及到系统的 Agent 策略的各个参数. 该方程收敛到稳态状态时的最短时间为短的 time-scale 的下确界.

从假设 1 - 假设 2 说明如下两点:

1) 在复杂管理系统中, 有足够多的主体在相互作用, 参与管理组织的运作. 这一系统由两个层次构成, 上层称之为 Super-Agent(局域), 下层称之为 Agent. 每一个 Super-Agent 中还有若干个不同的 Agent.

2) 同一局域内的 Agent 之间的相互作用关系是合作博弈, 而不同局域之间的相互作用关系是非合作博弈. 局域内合作博弈的目的是为了达到局域内 Agent 行为的协调化或同步, 从而提高整个局域的综合竞争能力. 而不同局域之间非合作博弈, 将使得各个局域之间产生竞争, 于是, 系统

行为之间存在差异, 这种差异使得系统不断的创新.

2 博弈模型: 确定型多局域世界图

本文将确定的拓扑结构抽象为一个确定型多局域世界图, 各个 Agent 之间的博弈关系抽象成一个多局域世界图中的随机微分博弈模型, 通过解析这一模型, 确定系统中 Agent 的最优策略.

将复杂管理系统中 Agent 相互作用之间的关系记为图 G , 设图 G 由 m 个子图构成(这些子图都应该属于密集图^[10], 在文献 [11] 的方法上实现分割), 子图 i 中的第 j 个 Agent 为 j_i , 则各个 Agent 之间相互作用的拓扑结构为其邻接矩阵 $G = (G_1, G_2, \dots, G_m)$.

下面分别讨论局域之内 Agent 的合作博弈、局域之间 Agent 的非合作博弈、以及 Agent 之间的合作 - 非合作博弈.

考虑系统中每个 Super-Agent 的状态约束, 得随机偏微分方程

$$dx_i(s) = f^i[s, x_i(s), \mu_i(s)] ds + \sigma_i[s, x_i(s)] dz_i(s), x_i(t_0) = x_i^0, j \neq i \quad (1)$$

其目标方程为

$$E_{t_0} \left\{ \int_{t_0}^T g^j[s, x_j(s), \mu_j(s)] \exp \left[- \int_{t_0}^s r(y) dy \right] ds + \exp \left[- \int_{t_0}^T r(y) dy \right] q^j(x_j(T)) \right\} \quad (2)$$

方程(2) 假定了行为的贴现因子函数为指数分布, 大多数管理系统都满足这一特征.

每个 Super-Agent 在参与管理活动时, 都有一个预期的目标(见方程(2)), 而这一目标实现的程度由这一 Super-Agent 携带的资源决定(见方程(1)). 在目标方程中, 可以得到如下信息: 每一 Super-Agent 追求的是在 time-scale $[t_0, T]$ 上收益最大, 其自身收益的大小由他资源的最佳投入量 x 、所使用的策略 u 以及当时的时间 t 直接相关; 每一个 Super-Agent 都具有目标实现的顺利程度, 在这里定义为贴现因子函数 $\exp \left[- \int_{t_0}^s r(y) dy \right] ds$. 在资源约束中, 可以得到如下信息: 每一个 Super-Agent 的资源不是独立的, 而是在与其他 Super-Agent 之间相互作用过程中在系统中流动, 资源量

的大小受到邻居资源大小及结构相互作用的影响; Super-Agent 在资源的投入上具有记忆性, 当投资亏损, 它会立刻调整策略, 减少投资量, 当投资赢利, 他将增加投资; 资源的变化受到前期资源的投资、Super-Agent 自身策略相互作用的影响, 这正好与其自身的目标形成一个反馈; Super-Agent 的资源变化受到外部环境及其邻居的相互作用, 表现出一定的随机性, 这一随机性主要表现在其漂移项 $\sigma_i [s, x_i(s)] dz_i(s)$ 中, 漂移项包括了其资源的波动性及一个 Weiner 过程, 具有一定的滤波功能, 这就使得 Super-Agent 具有对环境的预测功能。

同样的, 对于相应的 Agent 行为可以描述为

$$E_{t_0} \left\{ \int_{t_0}^T \sum_{j_i \in K_i} g^{j_i} [s, x_{j_i}(s), \mu_{j_i}(s)] \exp \left[- \int_{t_0}^s r(y) dy \right] ds + \sum_{j_i \in K_i} \exp \left[- \int_{t_0}^T r(y) dy \right] q^{j_i}(x_{j_i}(T)) \right\},$$

$$j_i \in K_i = \{1, 2, \dots, k_i\} \subseteq N_i \quad (3)$$

系统状态及其变化为

$$dx_{i_i}(s) = f^{j_i} [s, x_{j_i}(s), x_{k_i}(s), \mu_{j_i}(s)] ds + \sigma_{j_i} [s, x_{j_i}(s)] dz_{j_i}(s), \quad x_{j_i}(t_0) = x_{j_i}^0, \quad k_i \neq j_i \quad (4)$$

3 确定型多局域世界图中 Agent 行为均衡解 1: 局域内 Agent 的合作博弈

对于第 2 节描述的静态复杂管理系统中 Agent 之间的博弈, 需要确定以下两个重点: 1) 最优化每一个局域的收益; 2) 对于任意的局域, 需将获得的最优化收益进行合理的分配. 这两个特征将会使得该复杂管理系统具有长期持续的综合竞争能力。

按照第 1 节中的假设, 不同的 Super-Agent 之间的相互作用是非合作随机微分博弈驱动, 同一局域内的 Agent 之间的相互作用是以合作随机微分博弈驱动. 第 3 节和第 4 节将针对这两种不同的行为进行较为深刻的分析。

3.1 子系统的收益最大化

考虑任意一个局域 i 的合作随机微分博弈, 其局域的系统目标为

$$\max_{u_{N_i}} E_{t_0} \left\{ \int_{t_0}^T \sum_{j_i=1}^{n_i} g^{j_i} [s, x_{j_i}(s), \mu_{j_i}(s)] \times \exp \left[- \int_{t_0}^s r(y) dy \right] ds + \sum_{j_i=1}^{n_i} \exp \left[- \int_{t_0}^T r(y) dy \right] q^{j_i}(x_{j_i}(T)) \right\},$$

$$j_i \in N_i = \{1, 2, \dots, n_i\} \quad (5)$$

系统状态及其变化为

$$s. t. \quad dx_{i_i}(s) = f^{j_i} [s, x_{j_i}(s), x_{k_i}(s), \mu_{j_i}(s)] ds + \sigma_{j_i} [s, x_{j_i}(s)] dz_{j_i}(s), \quad x_{j_i}(t_0) = x_{j_i}^0, \quad k_i \neq j_i \quad (6)$$

子系统收益最大化问题, 有两个问题需要解决, 第一, 是确定局域的最优策略; 第二, 是在一个局域内将各个 Agent 的利益进行公平的分配^[10-14]. 这两个问题都是一个随机优化问题. 不妨记模型 (5) - (6) 的最优控制为 $\Gamma [N_i; t_0, x_0]$.

定理 1 给出了局域的最优策略, 解决了子系统收益最大化问题的第一个问题。

定理 1 对于模型 (5) - (6), 其最优控制 $\Gamma [N_i; t_0, x_0]$ 的最优策略集为 $\{u_{N_i}^*(t) = \psi_{N_i}^{t_0*}(t, x_{N_i})\}$. 如果存在一个连续可微函数 $W^{(t_0)N_i}(t, x_{N_i}) : [t_0, T] \times \prod_{j_i \in N_i} R^{m_{j_i}} \rightarrow R$ 满足偏微分方程

$$-W_t^{(t_0)N_i}(t, x_{N_i}) - \frac{1}{2} \sum_{h, \zeta=1}^m \Omega_{N_i}^{h\zeta}(t, x_{N_i}) W_{x^h x^\zeta}^{(t_0)N_i}(t, x_{N_i}) = \max_{u_{N_i}} \left\{ \sum_{j_i \in N_i} g^{j_i} [t, x_{j_i}, \mu_{j_i}] \exp \left[- \int_{t_0}^t r(y) dy \right] + \sum_{j_i \in N_i} W_{x_{j_i}}^{(t_0)N_i}(t, x_{N_i}) f_{j_i}^{N_i} [t, x, \mu] f_{j_i}^{N_i} [t, x_{k_i}, \mu_{j_i}] \right\}$$

其边界条件为

$$W^{(t_0)N_i}(T, x_{N_i}) = \sum_{j_i \in N_i} \exp \left[- \int_{t_0}^T r(y) dy \right] q^{j_i}(x_{j_i})$$

其中 $\Omega_{N_i}(t, x_{N_i})$ 是一个元素为 $\Omega_{j_i}(s, x_{j_i}(s))$ 的矩阵, $j_i \in N_i$, 该矩阵除了对角线元素外其他元素为 0.

$\Omega_{N_i}^{h\zeta}(t, x_{N_i})$ 是矩阵 $\Omega_{N_i}(t, x_{N_i})$ 的第 h_i 行第 ζ_i 列元素, $\frac{1}{2} \sum_{h, \zeta=1}^m \Omega_{N_i}^{h\zeta}(t, x_{N_i}) W_{x^h x^\zeta}^{(t_0)N_i}(t, x_{N_i})$ 被用来描述系统的随机性(下文亦如此)。

注释 1 1) 由于合作随机微分博弈模型 (5) - (6) 的最优解非常难获得, 需要将这一博弈分解成两部分来求解, 第一部分就是整个局域的收益最大化,

第二部分就是局域内各个 Agent 所分配的收益合理. 第一步主要描述的是集体理性约束条件, 本质上就是一个随机微分优化问题. 定理 1 将这一问题的最优解转换为一个 PDE 的解. 通过求解这一方程, 就可以知道, 该局域的最大收益, 以及对应的 Agent 的最优策略状态.

2) 所谓的策略, 其实就是各种资源在某一个时间点上的最优配置. 这些资源包括人、资金、商品、信息等等. 最优策略与时间有关系, 因此得到的最优策略应该是一个多维函数形成的轨迹.

3) 系统收益最优, 实际上是整体收益最大. 这时所得到的资源量等于局域内所有 Agent 的资源量的总和, 称之为集中决策.

4) 如果一个经济或管理组织, 需要最小的努力, 在最大程度上实现某一目标的价值, 这个组织就需要将自己的资源合理的分配, 达到一个最优结构. 这一最优结构的资源大小可通过定理 1 计算.

下面探讨第二个问题.

方程 (6) 中考虑了 Agent j_i 与在同一局域内的其他 Agent 之间的相互作用. 这一相互作用依赖于系统的拓扑结构, 当且仅当矩阵元素 $c_{j_i k_i} = 1$ 时, Agent j_i 与 Agent k_i 之间的相互作用才能体现出来, 才能形成合作关系, 获得多种合作联盟所导致收益公平分配下的最优策略.

同理, 可以确定局域内, Agent 之间相互作用而形成的各种联盟 K_i . 最优收益时对应的各个 Agent 的策略, 其对应的方程为

$$\begin{aligned} \max_{u_{K_i}} E_{t_0} \left\{ \int_{t_0}^T \sum_{j_i=1}^{n_i} g^{j_i} [s, x_{j_i}(s), \mu_{j_i}(s)] \times \right. \\ \left. \exp \left[- \int_{t_0}^s r(y) dy \right] ds + \right. \\ \left. \sum_{j_i=1}^{k_i} \exp \left[- \int_{t_0}^T r(y) dy \right] q^{j_i}(x_{j_i}(T)) \right\}, \\ j_i \in K_i = \{1, 2, \dots, k_i\} \subset \{1, 2, \dots, n_i\} = N_i \end{aligned} \quad (7)$$

$$\begin{aligned} \text{s. t. } dx_{j_i}(s) &= f^{j_i} [s, x_{j_i}(s), x_{k_i}(s)], \\ u_{j_i}(s) ds &+ \sigma_{j_i} [s, x_{j_i}(s)] dz_{j_i}(s), \end{aligned} \quad (8)$$

$$x_{j_i}(t_0) = x_{j_i}^0, \quad k_i \neq j_i$$

模型 (7) - (8), 可以得到最优策略.

定理 2 给出了在一确定局域内, 由于 Agent 之间相互作用而形成的各种联盟 K_i 最优收益时

的最优策略, 这就解决了子系统收益最大化问题的第二个问题.

定理 2 对于模型 (7) - (8), 其最优控制 $\Gamma [K_i; t_0, x_0]$ 的最优策略集为 $\{u_{K_i}^*(t) = \psi_{K_i}^{t_0^*}(t, x_{K_i})\}$. 如果存在一个连续可微函数 $W^{(t_0)K_i}(t, x_{K_i}) : [t_0, T] \times \prod_{j_i \in K_i} R^{m_{j_i}} \rightarrow R$, 满足偏微分方程

$$\begin{aligned} - W_t^{(t_0)K_i}(t, x_{K_i}) - \frac{1}{2} \sum_{h, \zeta=1}^m \Omega_{K_i}^{h\zeta}(t, x_{N_i}) W_{x_{h\zeta}}^{(t_0)K_i}(t, x_{K_i}) = \\ \max_{u_{K_i}} \left\{ \sum_{j_i \in K_i} g^{j_i} [t, x_{j_i}, \mu_{j_i}] \exp \left[- \int_{t_0}^t r(y) dy \right] + \right. \\ \left. \sum_{j_i \in K_i} W_{x_{j_i}}^{(t_0)K_i}(t, x_{K_i}) f_{j_i}^{K_i} [t, x, \mu] \right\} \end{aligned}$$

其边界条件为

$$W^{(t_0)K_i}(T, x_{K_i}) = \sum_{j_i \in K_i} \exp \left[- \int_{t_0}^T r(y) dy \right] q^{j_i}(x_{j_i}).$$

注释 2 1) 即使在同一个局域中, 也有可能随机形成若干个不同的联盟, 所有可能的联盟

总数量应该为 $\sum_{k=1}^n C_n^k$ 种.

2) 每一个 Agent 可以参加任意的一个联盟, 当 $k = 1$ 时, 这一联盟只有一个 Agent. 在不同的联盟中, Agent 可以获得不同的收益, Agent 通过综合比较, 确定自己能够获得最大收益的条件. 当然, 参加任何一个联盟与别的 Agent 进行合作时所得到的收益应该不低于这些最大收益的平均值.

3) 在一个经济或管理系统中, 局域中的参与者都可以参加任何一个联盟, 从而获得一定的收益. 获得的最大收益、最优的策略以及对应的联盟, 可以通过定理 2 获得. 与定理 1 一样, 最优策略与时间相关, 其轨迹是一个函数.

4) 局域中 Agent 参加联盟这一假设是必要的, 描述了一种虚拟的联盟形式. 通过定理 2, 可以知道 Agent 所有可能的收益与策略的集合 (局部最优条件) 所对应的一个决策 (分散决策).

定理 2 是 Agent 在局域内分配收益时的基础.

3.2 局域内收益的 Shapley 分配

对于局域内 Agent 收益的分配, 需要达到公平、合理、公正的效果, 为此, 提出以下条件: (1) 所分配的收益之和必须等于局域最大收益; (2) 对每一个 Agent, 参与合作所分配的收益应该不小

于不参与合作所得到的收益. 前者称作为集体理性, 后者称之为个体理性. 对局域内所有的 Agent, 他们都在力争获得更大的分配收益, 从而引发了合作博弈. 因此, 合作博弈的关键在于设计一种合理的分配机制, 使得每个 Agent 感受“合理”.

本文将采用动态 Shapley 值法^[15-21], 对任意局域内的 Agent 的收益进行分配, 并确定其对应的策略. 下面根据研究对象, 给出对应的条件.

条件 1

$$v^{(\tau)j_i}(\tau, x_{N_i}^{\tau*}) = \sum_{K_i \subseteq N_i} \frac{(k_i - 1)!(n_i - k_i)!}{n_i!} \times [W^{(\tau)K_i}(\tau, x_{K_i}^{\tau*}) - W^{(\tau)K_i \setminus j_i}(\tau, x_{K_i \setminus j_i}^{\tau*})], j_i \in N_i$$

$\tau \in [t_0, T]$ 及 $x_{N_i}^{\tau*} \in X_{N_i}^{\tau*}$ 其中

$$v^{(\tau)i}(\tau, x_{N_i}^{\tau*}) = [v^{(\tau)1}(\tau, x_{N_i}^{\tau*}) \quad v^{(\tau)2}(\tau, x_{N_i}^{\tau*}), \dots, v^{(\tau)n_i}(\tau, x_{N_i}^{\tau*})]$$

满足下列分配向量的基本属性:

- 1) $\sum_{j_i=1}^{n_i} v^{(\tau)j_i}(\tau, x_{N_i}^{\tau*}) = W^{(\tau)N_i}(\tau, x_{N_i}^{\tau*})$ 及
- 2) $v^{(\tau)j_i}(\tau, x_{N_i}^{\tau*}) \geq W^{(\tau)j_i}(\tau, x_{N_j}^{\tau*})$, 对于任意的 $j_i \in N_i$, 以及 $\tau \in [t_0, T]$.

其中条件中的 1) 保证了在博弈期间, $v^{(\tau)}(\tau, x_{N_i}^{\tau*})$ 满足 Pareto 最优以及集体理性. 2) $v^{(\tau)}(\tau, x_{N_i}^{\tau*})$ 保证了个体理性约束条件. $\mathcal{B}_{j_i}(s)$ 记做 Agent $j_i \in N_i$ 在时间 $s \in [t_0, T]$ 上的由 $v^{(t_0)j_i}(t_0, x_{N_i}^0)$ 决定的瞬时分配后的收益. 于是, 有条件 2.

条件 2 对于 $t \in [\tau, T]$ 及 $x_{N_i}^{\tau*} \in X_{N_i}^{\tau*}$,

$$\sum_{j_i=1}^{n_i} \mathcal{B}_{j_i}(s) = \sum_{j_i=1}^{n_i} g^{j_i}[s, x_{j_i}^{s*}, \psi_{N_i}^{(t_0)N_i^*}(s, x_{j_i}^{s*})]$$

显然, Shapley 值法可以满足条件 1 与 3.2 节. 按照动态 Shapley 值法以及合作博弈理论, 可以知道, 为了满足联盟收益的最大化, Agent 将在整个事件区间 $[t_0, T]$ 上采用策略变量 $\{\psi_{N_i}^{(t_0)N_i^*}(t, x_{N_i}^{t*})\}_{t=t_0}^T$, 相应的最优状态轨迹为 $\{x_{N_i}^*(t)\}_{t=t_0}^T$. 在时间 t_0 时刻状态为 $x_{N_i}^0$ 的系统中, 经过协调, 对于 Agent j_i 分享的收益为

$$v^{(t_0)j_i}(t_0, x_{N_i}^0) = \sum_{K_i \subseteq N_i} \frac{(k_i - 1)!(n_i - k_i)!}{n_i!} \times [W^{t_0 K_i}(t_0, x_{K_i}^0) - W^{t_0 K_i \setminus j_i}(t_0, x_{K_i \setminus j_i}^0)], j_i \in N_i$$

其中 $K_i \setminus j_i$ 是 $\{j_i\}$ 的补集.

以上分配仅仅适合于初始状态的分配. 随着

时间的推移, 系统中每个 Agent 的状态都发生了变化, 其策略将会随之而变, 这种分配就不再适合. 为此, 需要引入一个动态的分配机制以适应真实复杂管理系统的决策. 在本节中, 将条件 1 中确定 Shapley 值的原则给予外推, 构建一个暂静态补偿机制. 其过程如下: 首先, 构建一个分配过程, 使得分配方案满足条件 1. $\mathcal{B}_{j_i}(s)$ 表示 Agent $j_i \in N_j$ 在时间 $s \in [t_0, T]$ 由 $v^{(t_0)j_i}(t_0, x_{N_i}^0)$ 支配的收益. 特别的

$$v^{(t_0)j_i}(t_0, x_{N_i}^0) = \sum_{K_i \subseteq N_i} \frac{(k_i - 1)!(n_i - k_i)!}{n_i!} \times [W^{(t_0)K_i}(t_0, x_{K_i}^0) - W^{(t_0)K_i \setminus j_i}(t_0, x_{K_i \setminus j_i}^0)] = E_{t_0} \left\{ \int_{t_0}^T \mathcal{B}_{j_i}(s) \exp \left[- \int_{t_0}^s r(y) dy \right] ds + q^{j_i}(x_{j_i}^*(T)) \exp \left[- \int_{t_0}^T r(y) dy \right] \mid x_{N_i}(t_0) = x_{N_i}^0 \right\} = E_{t_0} \left\{ \int_{t_0}^T \mathcal{B}_{j_i}(s) \exp[-r(s - t_0)] ds + q^{j_i}(x_{j_i}^*(T)) \exp[-r(T - t_0)] \mid x_{N_i}(t_0) = x_{N_i}^0 \right\}, j_i \in N_i \quad (9)$$

同理, 对 $j_i \in N_i$ 和 $t \in [t_0, T]$, 用

$$v^{(t_0)j_i}(t_0, x_{N_i}^0) = E_{t_0} \left\{ \int_{t_0}^T \mathcal{B}_{j_i}(s) \exp[-r(s - t_0)] ds + q^{j_i}(x_{j_i}^*(T)) \exp[-r(T - t_0)] \mid x_{N_i}(t) = x_{N_i}^{t*} \right\} \quad (10)$$

表示合作收益在时间区间 $[t, T]$, Agent j_i 的当前价值. 该价值由时间 $t \in [t_0, T]$ 上的状态 $x_{N_i}^{t*} \in X_{N_i}^{t*}$ 给定.

显然, $v^{(t_0)j_i}(t, x_{N_i}^{t*})$ 能够满足条件 1 的一个必要条件是

$$\text{对于 } j_i \in N_i, t \in [t_0, T], x_{N_i}^{t*} \in X_{N_i}^{t*}, v^{(t_0)j_i}(t_0, x_{N_i}^{t*}) = v^{(t)j_i}(t, x_{N_i}^{t*}) \exp[-r(t - t_0)] \quad (11)$$

从上面可以看出, 能够同时满足 (9) - (10) 的 $v^{(t_0)j_i}(t, x_{N_i}^{t*})$ 的一个自然的选择是

$$v^{(t_0)j_i}(t, x_{N_i}^{t*}) = \sum_{K_i \subseteq N_i} \frac{(k_i - 1)!(n_i - k_i)!}{n_i!} \times [W^{(t_0)K_i}(t, x_{K_i}^{t*}) - W^{(t_0)K_i \setminus j_i}(t, x_{K_i \setminus j_i}^{t*})]$$

其中 $x_{N_i}^{t*} \in X_{N_i}^{t*}$.

为了实现这一构造 给出定理 3.

定理 3 Agent $j_i \in N_i$ 在时间 $\tau \in [t_0, T]$ 的 暂静态分配收益等于

$$\mathcal{B}_{j_i}(\tau) = - \sum_{K_i \subseteq N_i} \frac{(k_i - 1)!(n_i - k_i)!}{n_i!} \{ [W_t^{(\tau)K_i}(t, x_{K_i}^{\tau*}) |_{t=\tau}] - [W_t^{(\tau)K_i} \setminus j_i(t, x_{K_i}^{\tau*}) |_{t=\tau}] + [W_{x_{N_i}^{\tau*}}^{(\tau)K_i} \setminus j_i(t, x_{N_i}^{\tau*}) |_{t=\tau}] - [W_{x_{N_i}^{\tau*}}^{(\tau)K_i} \setminus j_i(t, x_{N_i}^{\tau*}) |_{t=\tau}]] f^{N_i}[\tau, x_{N_i}^{\tau*}, \psi_{j_i}^{(\tau)N_i}(\tau, x_{N_i}^{\tau*})] + \frac{1}{2} \sum_{h_i, \xi_i=1}^{n_i} \Omega_{K_i}^{h_i \xi_i}(\tau, x_{\tau}^*) [W_{x_{h_i \xi_i}^{\tau*}}^{(\tau)K_i} \setminus j_i(t, x_{\tau}^*) |_{t=\tau}] - \frac{1}{2} \sum_{h_i, \xi_i=1}^{n_i} \Omega_{K_i \setminus j_i}^{h_i \xi_i}(\tau, x_{\tau}^*) [W_{x_{h_i \xi_i}^{\tau*}}^{(\tau)K_i \setminus j_i} \setminus j_i(t, x_{\tau}^*) |_{t=\tau}] \}$$

其中 $x_{N_i}(\tau) = x_{N_i}^{\tau*} \in X_{N_i}^{\tau*}$ 将致使条件 1 得以实现.

(证明略)

根据以上分析 得到 1) Agent j_i 将自己的收益水平努力提高到 $g^{j_i}[s, x_{j_i}, \mu_{j_i}]$ 且 $q^{j_i}(x_{j_i})$ 正相关与收益 x_{j_i} . 2) Agent 能够与其他 Agent 形成联盟的收益函数中获得的关键收益将会导致

$$W^{(\tau)K}(\tau, x_K^{\tau*}) \geq W^{(\tau)L}(\tau, x_L^{\tau*}) + W^{(\tau)K \setminus L}(\tau, x_{K \setminus L}^{\tau*}),$$

$L \subset K \subseteq N$

其中 $K \setminus L$ 是 L 在 K 中的补集.

定理 3 给出了 Agent 的暂静态分配收益. 为使得收益在 Agent 中合理分配 定理 4 将进一步给出了满足条件 1 的 Agent 暂静态补偿收益.

注释 3 1) 当一个局域收益达到最大时 即定理 1 得以满足时 将使得整个局域的最大收益在局域内各个 Agent 中合理分配.

2) 对于任何一个 Agent 进入一个局域并能与其他 Agent 合作的条件是: 参加合作得到的收益不低于独立实现目标所得到的收益. 这要求每一个 Agent 努力发挥自己的核心竞争力 并在定理 3 设计的机制下 通力合作 从而不仅使得整体的收益最大 而且使得自己的收益大于独立运营时得到的收益. 体现了合作随机微分博弈的意义.

3) 与定理 1.3.2 节一致 分配到的收益是关于时间的一个映射 也是动态的 并且形成了一个最优分配的轨迹.

定理 4 Agent $j_i \in N_i$ 在时间 $\tau \in [t_0, T]$ 能够导

致条件 1 的暂静态补偿收益为

$$\mathcal{B}_{j_i}(\tau) = - \sum_{K_i \subseteq N_i} \frac{(k_i - 1)!(n_i - k_i)!}{n_i!} \times \{ [W_t^{(\tau)K_i}(t, x_{K_i}^{\tau*}) |_{t=\tau}] - [W_t^{(\tau)K_i} \setminus j_i(t, x_{K_i}^{\tau*}) |_{t=\tau}] + \sum_{i_j \in K_i} [W_{x_{j_i}^{\tau*}}^{(\tau)K_i}(t, x_{K_i}^{\tau*}) |_{t=\tau}] f_i^{N_i}[\tau, x_{N_i}^{\tau*}, \psi_{j_i}^{(\tau)N_i}(\tau, x_{N_i}^{\tau*})] - \sum_{i_h \in K_i \setminus j_i} [W_{x_{i_h}^{\tau*}}^{(\tau)K_i \setminus j_i}(t, x_{K_i \setminus j_i}^{\tau*}) |_{t=\tau}] f_h^{N_i}[\tau, x_{N_i}^{\tau*}, \psi_{i_h}^{(\tau)N_i}(\tau, x_{N_i}^{\tau*})] + \frac{1}{2} \sum_{h_i, \xi_i=1}^{n_i} \Omega_{K_i}^{h_i \xi_i}(\tau, x_{\tau}^*) [W_{x_{h_i \xi_i}^{\tau*}}^{(\tau)K_i}(t, x_{\tau}^*) |_{t=\tau}] - \frac{1}{2} \sum_{h_i, \xi_i=1}^{n_i} \Omega_{K_i \setminus j_i}^{h_i \xi_i}(\tau, x_{\tau}^*) [W_{x_{h_i \xi_i}^{\tau*}}^{(\tau)K_i \setminus j_i} \setminus j_i(t, x_{\tau}^*) |_{t=\tau}] \}$$

其中 $x_{N_i}(\tau) = x_{N_i}^{\tau*} \in X_{N_i}^{\tau*}$.

(证明略)

根据以上分析 得到 1) 向量 $\mathcal{B}_{j_i}(\tau)$ 作为一个均衡的暂静态补偿 能够保证 Shapley 值的分配在整个博弈区间中得以实现. 2) Agent j_i 在时间 τ 上的瞬间收益 $\mathcal{B}_{j_i}(\tau)$ 是以当前状态 $x_{N_i}^{\tau*}$ 和当前时间 τ 为条件的 因此可以将 $\mathcal{B}_{j_i}(\tau)$ 表达为 $\mathcal{B}_{j_i}(\tau, x_{N_i}^{\tau*})$, 即 Agent j_i 的瞬时的收益 $\mathcal{B}_{j_i}(\tau, x_{N_i}^{\tau*})$ 可以导致联盟的动态稳定解.

注释 4 定理 4 描述了当系统中各个 Agent 的约束有差异的情况 其他的管理方面的解释与定理 3 相同.

定理 3 与 4 描述了具有同质指数分布这一特征的贴现因子函数 在系统中、同一局域内各个 Agent 利益分配及补偿的情况.

4 确定型多局域世界图中 Agent 行为均衡解 2: 局域之间 Agent 的非合作博弈

4.1 非合作博弈下 Super-Agent 的均衡策略

本文将每一个局域抽象成为一个 Super-Agent , 假设整个系统可以分解成 m 个局域. 按照前面的假设 1 - 假设 2 , Super-Agent 之间的相互作用是非合作随机微分博弈 Γ , 本文将构造非合作随机

微分博弈模型.

对于任意的 Super-Agent i , 其目标可以确定为

$$E_{t_0} \left\{ \int_{t_0}^T g^i [s, x_i(s), \mu_i(s)] \exp \left[- \int_{t_0}^s r(y) dy \right] ds + \exp \left[- \int_{t_0}^T r(y) dy \right] q^i(x_i(T)) \right\}$$

$$i \in M \tag{12}$$

对应的约束条件为

$$dx_i(s) = f^i [s, x_i(s), \mu_i(s)] ds + \sigma_i [s, x_i(s)] dz_i(s),$$

$$x_i(t_0) = x_i^0 \tag{13}$$

对于每一个 Super-Agent, 他们之间的非合作随机微分博弈应该存在一个 Nash 均衡点, 每个 Super-Agent 都会有自己的 Nash 策略, 及其对应的收益.

引理 1 策略 n 元标量 $\{u_i^*(s) = \phi_i^*(t, x) \in U^i, i \in N\}$ 给出了博弈 (12) - (13) 的反馈 Nash 均衡解, 则存在一个连续的微分函数 $V^i(t, x) : [t_0, T] \times R^m \rightarrow R, i \in N$, 满足下列偏微分方程组:

$$-V_i^i(t, x) - \frac{1}{2} \sum_{h=1}^n \Omega^{hh}(t, x) V_{x_h x_h}^i =$$

$$\max_{u_i} \{ g^i [t, x, \phi_1^*(t, x), \phi_2^*(t, x), \dots, \phi_{i-1}^*(t, x),$$

$$u_i(t, x), \phi_{i+1}^*(t, x), \dots, \phi_n^*(t, x)] +$$

$$V_x^i(t, x) f [t, x, \phi_1^*(t, x), \phi_2^*(t, x), \dots, \phi_{i-1}^*(t, x),$$

$$u_i(t, x), \phi_{i+1}^*(t, x), \dots, \phi_n^*(t, x)] \}$$

$$=$$

$$g^i [t, x, \phi_1^*(t, x), \dots, \phi_n^*(t, x)] +$$

$$V_x^i(t, x) f [t, x, \phi_1^*(t, x), \dots, \phi_n^*(t, x)]$$

$$V^i(T, x) = q^i(x), i \in N$$

(证明略)

4.2 Super-Agent 的非合作与 Agent 的合作耦合

第 3 节和第 4.1 节分别给出了局域内的 Agent 合作博弈的最优策略, 以及 Super-Agent 非合作博弈的 Nash 均衡策略. 第 4.1 节中, 各个 Super-Agent 达到均衡状态时的局域获得的收益是 $V^i(t, x) : [t_0, T] \times R^m \rightarrow R$ 其中 $i \in M$. 显然这一收益并非将某一个 Super-Agent 孤立地进行优化之后获得的收益, 也就是并非 $W^i(t, x) : [t_0, T] \times R^m \rightarrow R$ 其中 $i \in M$. 但是在第 3 节的分析中, Agent 的策略及收益来自于 $W^i(t, x) : [t_0, T] \times R^m \rightarrow R$ 并

非是 $V^i(t, x) : [t_0, T] \times R^m \rightarrow R$, 即非合作与合作存在偏差. 因此必须纠正这一偏差, 称之为非合作与合作的耦合.

非合作与合作的耦合, 实际上是非合作博弈 $\Gamma(s, x, \mu)$ 的 Nash 均衡解 $V(s, x, \mu)$ 和最优化问题 $\varpi(s, x, \mu)$ 的最优解 $W(s, x, \mu)$ 之间的关系问题. 某一 Super-Agent 的 Nash 均衡解就是在其他 Super-Agent 收益都得到最优化的前提之下自身收益最大化对应的最优解; 各个 Super-Agent 的 Nash 最优解就是独立考虑某一 Super-Agent 最优化问题的可行解. 这二者之间存在着很强的关系. 本文采用变分理论寻找这一关系, 试图构造一个非线性算子, 来确定他们二者之间的联系, 引入陈光亚的一个结论^[22].

命题 1 设 (X, d) 是完备尺度空间, (Y, C) 是 Hausdorff 序拓扑空间且 $c^0 \in \text{int } C$. 设 $f: X \rightarrow Y$ 是向量值映射且 C -下有界. 设对给定的 $\varepsilon > 0$ 和每一个 $x \in X$ 集合

$$\{y \in X : f(y) - f(x) + \sqrt{\varepsilon} d(x, y) c^0 \leq c^0\}$$

是闭的. 则对每个 $x^0 \in X$ 使得 $f(x^0) \in \min_{\varepsilon c^0}(f(X))$, 存在点 $x_\varepsilon \in X$ 使得

- 1) $f(x_\varepsilon) \leq_C f(x^0)$
- 2) $d(x^0, x_\varepsilon) \leq \sqrt{\varepsilon}$
- 3) $f_{\varepsilon c^0}(x_\varepsilon) \in \min_C(f_{\varepsilon c^0}(X))$

其中 $f_{\varepsilon c^0}(x) = f(x) + \sqrt{\varepsilon} d(x, x^0) c^0, \forall x \in X$.

注意 $\min_{\varepsilon c^0}(A)$ 是集合 A 的 ε -有效点, $\min(A)$ 是 A 的有效点.

按照变分理论, 构造一个非线性算子 $\vartheta(s, x, u)$, 使得将非合作博弈 $\Gamma(s, x, \mu)$ 的 Nash 均衡解 $V(s, x, \mu)$ 和最优化问题 $\varpi(s, x, \mu)$ 的最优解 $W(s, x, \mu)$ 形成一个强联系: $V(s, x, \mu) = \vartheta(s, x, u) W(s, x, \mu) + o(h(s, x, \mu))$. 对于同质的指数分布为贴现因子函数系统有:

定理 5 存在使得 $V(s, x, \mu) = \varphi(W(s, x, u))$ 成立的函数 $\varphi(x)$, 对于贴现因子函数满足同质指数分布的经济与管理复杂自适应系统, 满足

如果 Γ 为一个无限维反馈博弈, 则 $\varphi(t, x) = \exp[-r(t - t_0)] x$;

如果 Γ 为一个有限维反馈博弈, 则 $\varphi(t, x) =$

$o(\exp[-r[t-t_0]]x^n)$,且 $0 < q < 1$.

(证明略)

定理5 将局域内 Agent 的合作随机微分博弈与局域 Super-Agent 之间的非合作随机微分博弈进行了耦合. 通过耦合便可以直接将局域最优化的结果修订为局域 Super-Agent 之间的非合作随机微分博弈的 Nash 均衡解. 将其在局域内进行分配, 从而实现复杂管理系统(多局域世界图中合作-非合作博弈为相互作用机制) Agent 的最优策略及收益.

注释5 1) 通过定理5, 可以知道在合作与非合作随机微分博弈混合作用下, 系统中任何一个 Agent 的策略与对应的最大收益的表达式.

2) 定理5 告诉人们, 系统中任意一个 Agent 的最优策略与最大收益依赖于以下特征: 局域内合作随机微分博弈的收益与策略、实现目标时的折扣因子、目标函数以及自身资源的动力学方程, 也就是任何一个 Agent 的最优策略与最大收益依赖于自身的特点.

3) 从管理意义上讲, 通过调查可以得到每一个 Agent 资源的行为规则、目标函数以及初始条件, 于是可以通过定理1~定理5 便可以确定出该 Agent 的最优策略及其最大收益, 以指导整个管理系统的决策.

5 最优策略的稳定性

随机微分博弈是一个动态博弈, 在系统随着 Agent 之间的相互作用, 以及受外部环境干扰时, 将会发生动态变化, 当且仅当系统中的 Agent 行为(策略)收敛到一个比较恒定的值时, 系统趋于稳定, 行为才能被称之为“确定”的. 这时只要出现“不和谐”的因素, 将会使得整个系统发生相变.

在第3节和第4节对优化问题、博弈问题求解的过程中, 笔者将其转化为一系列 Bellman 簇的方程, 显然, 这些方程解的收敛性直接决定了博弈解的收敛性, 因此, Agent 的相互作用经过某一段特定长的时间之后, 其策略趋于一个稳定的值.

因为收敛特性由描述系统特性的 Agent 行为的约束条件, 即状态动力学方程所对应的偏随机微分方程的收敛特性所决定^[24]. 因此可以将问题

转换为对 Agent 状态动力学方程吸引子的解析.

对于一个一般系统

$$d[x(t) - G(x_i)] = f(t, x_i) dt + g(t, x_i) dW(t) \tag{14}$$

其中 $G: C[-\tau, 0]R^n \rightarrow R^n$ $f: R_+ \times C[-\tau, 0]R^n \rightarrow R^n$ $g: R_+ \times C[-\tau, 0]R^n \rightarrow R^{n \times m}$ 均为定义在概率空间上的 Borel 可测函数, $W(t)$ 是 m 维的 Wiener 过程. $x_i = x(t+s)$ ($-\tau \leq s \leq 0$) (H) 给定的初始值 $\xi \in C_{F_0}^b([-\tau, 0]R^n)$, 方程(14) 在 $t \in [-\tau, \infty)$ 上有唯一解 $x(t, \xi)$.

定义1 严格正函数 $\psi(t)$ 为 ψ -函数, 如果满足 $\psi(0) = 1$ $\psi(\infty) = 0$,且对任意 $t \geq -\tau$, $\psi'(t) < 0$.

令 $\psi(t)$ 为一个既定的 ψ -函数, 令 $\psi_1(t) = \frac{\psi'(t)}{\psi(t)}$, 设

$$\begin{aligned} \gamma &:= \sup_{t \geq 0} |\psi(t)| < \infty, \\ \rho &:= \sup_{t \geq 0} \frac{\psi(t-\tau)}{\psi(t)} < \infty \end{aligned} \tag{15}$$

$$\begin{cases} h_\alpha(t) = \psi^{-\alpha}(t) x(t) \\ \tilde{h}_\alpha(t) = \psi^{-\alpha}(t) \tilde{x}(t) \\ \tilde{x}(t) = x(t) - G(x_i) \end{cases} \tag{16}$$

其中 $x(t) = x(t, \xi)$ 是方程(14) 的解.

引理2 不妨设存在常数 $\kappa \in (0, 1)$ 使得

$$|G(\varphi)| \leq \kappa \|\varphi\| \tag{17}$$

对于所有的 $\varphi \in C[-\tau, 0]R^n$. 令

$$\alpha \in \left[0, \frac{\ln \kappa^{-1}}{\ln \rho}\right] \text{ 及 } \kappa_1 = \kappa \rho^\alpha, \text{ 则对于方程}$$

(14) 给定的任意解 $x(t, \xi)$, 函数 $h_\alpha(t)$ 和 $\tilde{h}_\alpha(t)$ 满足

$$\sup_{-\tau \leq s \leq t} |h_\alpha(s)| \leq \frac{1}{1-\kappa_1} [\|\xi\| \vee \sup_{-\tau \leq s \leq t} |\tilde{h}_\alpha(s)|], \tag{18} \\ t \geq 0$$

下面定义其对应的吸引子.

定义2 令 Φ 是 R^n 的非空子集, 不妨设存在 $\alpha \geq 0$, 对于任意的初始值 $\xi \in C_{F_0}^b([-\tau, 0]R^n)$, $x(t, \xi)$ 为随机微分方程^[27]

$$d[x(t) - G(x_i)] = f(t, x) dt + g(t, x_i) dW(t)$$

的解. 称 $C_{F_0}^b([-\tau, 0]R^n)$ 以概率1弱吸引于点 Φ , 如果 $h_\alpha(t, \xi)$ 满足

$$\liminf_{t \rightarrow \infty} d(h_\alpha(t, \xi), \Phi) = 0, \text{ a. s.} \quad (19)$$

称 Φ 以概率 1 收敛于点 $\xi \in C_{\mathcal{C}_0}^b([- \tau, \rho] R^n)$, 如果

$$\lim_{t \rightarrow \infty} d(h_\alpha(t, \xi), \Phi) = 0, \text{ a. s.} \quad (20)$$

引理 3 设 (H) 成立, 假定存在函数 $V(t, x) \in C^1(R_+ \times R^n, R_+)$, $\gamma(t) \in L^1(R_+, R_+)$, $\mu_1(x), u_2(x) \in C(R^n, R_+)$ 及 $\zeta: [- \tau, \rho] \rightarrow R_+$, 使得 $\int_{-\tau}^0 \zeta(s) ds = 1$. 同时存在常数 $p \geq 1$ 和 $c > 0$, 使得

$$L[\psi^{-p\alpha}(t) V(t, \varphi)] \leq \gamma(t) - u_1(\psi^{-\alpha}(t) \varphi(0)) + \int_{-\tau}^0 \zeta(\theta) u_2(\psi^{-\alpha}(t + \theta) \varphi(\theta)) d\theta \quad (21)$$

对于所有 $(t, \varphi) \in R_+ \times C([- \tau, \rho] R^n)$. 假设 $u(x) = u_1(x) - u_2(x) \geq 0$ 对所有 $x \in R^n$ 有

如果 $\alpha = 0$,

$$\liminf_{t \geq 0, |x| \rightarrow \infty} V(t, x) = \infty \quad (22)$$

如果 $\alpha > 0$,

$$c |x|^p \leq V(t, x) \quad (23)$$

则 $\ker(u) = \{x(t) \in R^n : u(h_\alpha(t)) = 0\} \neq \emptyset$, 即

$$\liminf_{t \rightarrow \infty} d(h_\alpha(t), \ker(u)) = 0, \text{ a. s.} \quad (24)$$

引理 3 表明 $h_\alpha(t)$ 将以概率 1 无限次访问 $\ker(u)$ 的邻域, 也就是说, $\ker(u)$ 吸引 $h_\alpha(t)$ 无限次, 致使一些学者认为 $\ker(u)$ 是 $h_\alpha(t)$ 的弱吸引子. 然而, 这并不保证 $h_\alpha(t)$ 最终被 $\ker(u)$ 吸引.

(证明略. 强收敛性见定理 6).

定理 6 假设存在常数 $\kappa \in (0, 1)$ 使得条件 (17) 满足, 又假定存在另外一个常数 $\alpha, \rho \leq \alpha \leq -\frac{\ln \kappa}{\ln \rho}$ 使得对于任意的 $\gamma > 0$

$$k_r := \sup_{t \geq 0, \|\varphi\| \leq r\psi^\alpha(t-\tau)} \psi^{-\alpha}(t) [|f(t, \varphi)| \vee |g(t, \varphi)|] < \infty$$

在引理 3 条件下, 假定

$$\lim_{t \rightarrow \infty} u(\tilde{h}_\alpha(t)) = 0 \text{ a. s.} \Rightarrow \lim_{t \rightarrow \infty} u(h_\alpha(t)) = 0 \text{ a. s.}$$

$$E \int_0^\infty u(h_\alpha(t)) dt < \infty \Rightarrow E \int_0^\infty u(\tilde{h}_\alpha(t)) dt < \infty,$$

则 $\ker(u) = \{x(t) \in R^n : u(h_\alpha(t)) = 0\} \neq \emptyset$, 并且 $C_{\mathcal{C}_0}^b([- \tau, \rho] R^n)$ 以概率 1 收敛到 $\ker(u)$, 即 $\lim_{t \rightarrow \infty} d(h_\alpha(t), \ker(u)) = 0$ a. s.

(证明略).

以联盟 $K_i \subseteq N_i$ 为例描述同质指数分布为贴现因子的对应的方程, 具体结果参见定理 1. 当且仅当定理 1 对应的 Bellman 方程的解收敛时, 系统内 Agent 的 Pareto-Nash 最优策略才能收敛到其吸引子. 下面, 确定这三个系统对应的 Lyapunov 函数 $\mathcal{L}(W(t, x))$.

$$\begin{aligned} \mathcal{L}(W(t, x)) &= W_t^{(t_0) K_i}(t, x_{K_i}) + \\ &\frac{1}{2} \sum_{h, \xi=1}^m \Omega_{K_i}^{h, \xi}(t, x_{N_i}) W_{x_{h, \xi}}^{(t_0) K_i}(t, x_{K_i}) + \\ &\max \left\{ \sum_{j_i \in K_i} g^{j_i}[t, x_{j_i}, \mu_{j_i}] \exp \left[- \int_{t_0}^t r(\gamma) dy \right] + \right. \\ &\left. \sum_{j_i \in K_i} W_{x_{j_i}}^{(t_0) K_i}(t, x_{K_i}) f_{j_i}^{K_i}[t, x, \mu] \right\} \end{aligned} \quad (25)$$

令 $\psi(t) = \exp[-r(t - t_0)]$, 由式 (25) 知道

$$L[\psi^{-p\alpha}(t) V(t, \varphi)] \leq \gamma(t) - u_1(\psi^{-\alpha}(t) \varphi(0)) + \int_{-\tau}^0 \zeta(\theta) u_2(\psi^{-\alpha}(t + \theta) \varphi(\theta)) d\theta = \gamma(t) - u_1(\psi^{-\alpha}(t) \varphi(0)) + E(u_2(\psi^{-\alpha}(t) \varphi(t))) \quad (26)$$

不妨令 $p = 1$, $\varphi(t) = x(t)$ 根据 $x(t)$ 的特性, 可知

$$L[\psi^{-\alpha}(t) W(t, x)] \leq \gamma(t) - [u_1(h_\alpha(t, x)) - u_2(h_\alpha(t, x))] = \gamma(t) - u(h_\alpha(t, x))$$

按照式 (24) 以及定理 5, 易知, 存在一些令 $u(h_\alpha(t, x)) = 0$ 的 $x(t)$, 使得 $\lim_{t \rightarrow \infty} d(h_\alpha(t), \ker(u)) = 0$. 联立式 (26), 易知, 满足这一条件的解必然是方程 (26) 的解, 从而, 方程 (26) 也就是系统中 Agent 的行为收敛到最优策略 $(t, x_{K_i}^*)$.

这种情况下, 系统最优策略的收敛如图 1 所示.

这就是说, 在最优控制综合函数的作用下, 系统从任意初态开始, 以及遇到任意的扰动, 它都以最优的方式运行至终端 $x(T)$, 即获得的最优策略能够收敛于吸引子.

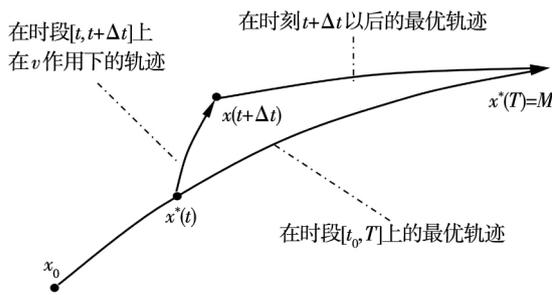


图1 随机扰动下系统中 Agent 最优策略的稳定性
 Fig. 1 The stability of Agent's optimal strategy with stochastic disturbance

6 结束语

本文研究基于人的行为的复杂管理系统. 在短的 time-scale 中, 如何确定在一个确定的系统拓扑结构下 Agent 的最优策略. 提出了一个确定型多局域世界图博弈模型, 并对确定型多局域世界图, 分两种情况讨论 Agent 行为均衡解: (1) 局域内 Agent 的合作博弈; (2) 局域之间的非合作博弈. 分别获得局域内的 Agent 基于合作博弈的最优策略, 基于非合作博弈的 Nash 均衡策略. 并将

局域内 Agent 的合作随机微分博弈与局域 Super-agent 之间的非合作随机微分博弈进行了耦合, 将收益在局域内合理分配. 最后讨论了最优策略的稳定性. 本文与以往研究不同之处是提出了在短的 time-scale 中, 一个确定的系统拓扑结构下 Agent 的最优策略确定的解析方法.

本文对在短的 time-scale 中 Agent 的最优策略进行了分析, 这只是经济与管理复杂自适应系统的开端. 正如在假设中提出的那样, 可以看出, 本文所研究的经济与管理复杂系统是从时间与空间两个维度按照不同粒度对系统进行分割的, 即从空间上按照“物以类聚、人以群分”的思想分割成不同的子系统(从大的空间-scale 看, 小的空间 scale 被压缩成一个个体); 从时间上按照系统结构的固定与变化将系统发展分割成短的 time-scale 和长的 time-scale 两种(从长的 time-scale 看, 短的 time-scale 被压缩成一个时间点). 由于篇幅限制, 本文并没有针对长 time-scale 中 Agent 的最优策略演化特征进行分析, 关于在长 time-scale 中变化系统结构下 Agent 最优策略的分布特征, 将在后续的文章中给予翔实的分析.

参考文献:

[1] Ke Hu, Tao Hu, Yi Tang. Cascade defense via control of the fluxes in complex networks [J]. J. Stat. Phys., 2010, 141: 555 - 565.

[2] 杨孟, 傅新楚, 吴庆初. 复杂网络上带传播媒介 SIS 模型的全局稳定性 [J]. 系统工程学报, 2010, 25(6): 767 - 773.

Yang Meng, Fu Xinchu, Wu Qingchu. Global stability of SIS epidemic model with infective medium on complex networks [J]. Journal of Systems Engineering, 2010, 25(6): 767 - 773. (in Chinese)

[3] Gao Z, Kong D, Gao C. Modeling and control of complex dynamic systems: Applied mathematical aspects [J]. Journal of Applied Mathematics, 2012, 2012(4): 1 - 18.

[4] Sethna J P. Entropy, Order Parameters, and Complexity [M]. Oxford: Oxford University Press, 2006.

[5] Cao W, Chen G, Chen X. Optimal tracking agent: A new framework of reinforcement learning for multi-agent systems [J]. Concurrency and Computation: Practice and Experience, 2013, 25: 2002 - 2015.

[6] 蒋国银, 胡斌. 集成博弈和多智能体的人群工作互动行为研究 [J]. 管理科学学报, 2011, 14(2): 29 - 41.

Jiang Guoyin, Hu Bin. Study on interaction between group and work based on game and multi-agent [J]. Journal of Management Sciences in China, 2011, 14(2): 29 - 41. (in Chinese)

[7] 龚日朝. 基于秩依期望效用理论的鹰鸽博弈均衡解分析 [J]. 管理科学学报, 2012, 15(9): 35 - 45.

Gong Rizhao. Nash equilibrium of hawk-dove game based on rank-dependent expected utility theory [J]. Journal of Management Sciences in China, 2012, 15(9): 35 - 45. (in Chinese)

[8] 王先甲, 全吉, 刘伟兵. 有限理性下的演化博弈与合作机制研究 [J]. 系统工程理论与实践, 2011, 30(增刊1): 82 - 93.

- Wang Xianjia, Quan Ji, Liu Weibing. Study on evolutionary games and cooperation mechanism within the framework of bounded rationality [J]. *Systems Engineering: Theory & Practice*, 2011, 30(Sup1): 82–93. (in Chinese)
- [9] Newman M E J. The structure and function of complex networks [J]. *SIAM Review*, 2003, 45(2): 167–256.
- [10] 陶少华, 张向群. 复杂网络自相似特征演化模型研究 [J]. *计算机工程*, 2012, 38(1): 197–199.
Tao Shaohua, Zhang Xiangqun. Research on self-similarity characteristic evolution model of complex network [J]. *Computer Engineering*, 2012, 38(1): 197–199. (in Chinese)
- [11] Nepusz T, Negyessy L. Reconstructing cortical networks: Case of directed graphs with high level of reciprocity [M]. In: BelaBollobas, Robert Kozma, DeasoMiklos. *Handbook of large-scale random networks*. Hungary: Springer, 2008: 325–368.
- [12] Cajueiro D O, Andrade R F S. Controlling self-organized criticality in complex networks [J]. *The European Physical Journal B*, 2010, 77: 291–296.
- [13] 王 龙, 伏 锋, 陈小杰, 等. 复杂网络上的群体决策 [J]. *智能系统学报*, 2008, 3(2): 95–107.
Wang Long, Fu Feng, Chen Xiaojie, et al. Collective decision-making over complex networks [J]. *CAAI Transactions on Intelligent Systems*, 2008, 3(2): 95–107. (in Chinese)
- [14] David Applebaum. *Lévy Processes and Stochastic Calculus* [M]. Cambridge: Cambridge University Press, 2009.
- [15] Alòs E, León J A, Vives J. On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility [J]. *Finance and Stochastics*, 2007, 11(4): 571–589.
- [16] 杜晓君, 马大明, 张 吉. 基于进化博弈的专利联盟形成研究 [J]. *管理科学*, 2010, 23(2): 38–44.
Du Xiaojun, Ma Daming, Zhang Ji. Research on patent pool coalition based on evolutionary game [J]. *Journal of Management Science*, 2010, 23(2): 38–44. (in Chinese)
- [17] 闫 安, 达庆利, 裴 凤. 多个企业同时博弈的动态古诺模型的长期产量解研究 [J]. *管理工程学报*, 2013, 27(1): 94–98.
Yan An, Da Qingli, Pei Feng. A study on the long-term produce solutions of dynamical model with multi-firms simultaneous-move game [J]. *Journal of Industrial Engineering/Engineering Management*, 2013, 27(1): 94–98. (in Chinese)
- [18] Szikora P. Introduction into the literature of cooperative game theory with special emphasis on dynamic games and the core [J]. *Proceedings-11th International Conference on Management, Enterprise and Benchmarking (MEB 2013)*, 2013.
- [19] Yan Wu, Jin Yaochu, Liu Xiaoxiong. A directed search strategy for evolutionary dynamic multiobjective optimization [J]. *Methodologies And Application*, 2015, 19(11): 3221–3235.
- [20] Javier Martínez-de-Albéniz F, Rafels C. Cooperative assignment games with the inverse Monge property [J]. *Discrete Applied Mathematics*, 2014, 162(1): 42–50.
- [21] Friesz T L. *Dynamic optimization and differential games* [M]. New York: Springer, 2010.
- [22] 陈光亚. 优化与均衡的等价性 [J]. *系统科学与数学*, 2009, 29(11): 1441–1446.
Chen Guangya. Equivalency between optima and equilibria [J]. *Journal of Systems Science & Mathematical Sciences*, 2009, 29(11): 1441–1446. (in Chinese)
- [23] 朱 亮, 韩定定. 动态复杂网络的同步拓扑演化 [J]. *计算机应用*, 2012, 32(2): 330–334, 339.
Zhu Liang, Han Dingding. Topological evolution on synchronization of dynamic complex networks [J]. *Journal of Computer Applications*, 2012, 32(2): 330–334, 339. (in Chinese)
- [24] Bollobás B. Percolation, Connectivity, Converge and Coloring [M]. In: BélaBollobas, Robert Kozma, PezsóMiklós. *Handbook of Large-Scale Random Networks*, Hungary: Springer, 2008: 117–142.
- [25] Bollobás B, Riordan O. Random graphs and branching process [M]. In: BélaBollobas, Robert Kozma, PezsóMiklós. *Handbook of Large-Scale Random Networks*. Hungary: Springer, 2008: 15–116.
- [26] 魏立峰, 陈 丽. 一类带松弛控制的 Hamilton-Jacobi-Bellman 方程的粘性解 [J]. *山东大学学报(理学版)*, 2009, 44(12): 1–5.
Wei Lifeng, Chen Li. Viscosity solutions of Hamilton-Jacobi-Bellman for the relaxed control [J]. *Journal of Shandong University (Natural Science)*, 2009, 44(12): 1–5. (in Chinese)

- [27] Cen Liqun, Dai Weixing, Hu Shigeng. Attraction and stability for neutral stochastic functional differential equations [J]. Wuhan University Journal of Natural Science, 2009, 14(3): 205–209.
- [28] Samothrakis S, Lucas S, Runarsson T P, et al. Coevolving game-playing agents: Measuring performance and intransitivities [J]. Evolutionary Computation, IEEE Transactions on, 2013, 17(2): 213–226.
- [29] 郭 英. 随机 Volterra 积分方程相容解的稳定性 [J]. 应用数学学报, 2011, 3(2): 331–340.
Guo Ying. Stability of adapted solutions of stochastic volterra-integro-equations [J]. Acta Mathematicae Applicatae Sinica, 2011, 3(2): 331–340. (in Chinese)

Optimal strategies of agents in deterministic multi-local-worlds graph

ZHENG Zhan¹, ZHENG Xiao-jing^{2*}, XU Xu-song³

1. Wuhan Textile University, Wuhan 430073, China;
2. Harbin University of Commerce, Harbin 150076, China;
3. Wuhan University, Wuhan 430072, China

Abstract: In a short time-scale, the optimal strategy of Agent for complex management systems is determined by using the analytic method in a deterministic system topology structure. A game model in the deterministic multi-local-worlds graph is constructed and the equilibrium solutions of Agents behavior are discussed for deterministic Multi-Local-Worlds graphs in two cases, which are the cooperative game between agents in the same local-world and non-cooperative game between agents in different Local-Worlds; The two different solutions with corresponding game models are coupled together to get the optimal strategies of Agents and the stability of the optimal strategy is discussed at the end of paper.

Key words: deterministic; multi-local-worlds graph; optimal strategies of agents; stochastic differential game