

基于交互的 Agent 行为选择动态性分析

江道平, 班晓娟, 尹怡欣

(北京科技大学信息工程学院, 北京 100083)

摘要:在动态不可预知环境下, Agent 行为选择过程表现出动态性。该文基于人工生命和 MAS 框架, 建立 Agent 行为选择动态系统模型, 运用动态系统观点研究 Agent 感知、认知交互和适应, 分析 Agent 行为选择适应外部环境和满足内部需求的动态平衡过程。在仿真试验中, 通过 21 个 Agent 局部感知和交互, 实现了全局平衡、稳定的群体运动行为。

关键词:感知; 认知; 行为选择; 动态性分析

Dynamic Analysis on Agents Action Selection Based on Interaction

JIANG Dao-ping, BAN Xiao-juan, YIN Yi-xin

(School of Information Engineering, University of Science & Technology Beijing, Beijing 100083)

【Abstract】Under the dynamical and unpredictable environment, Agents action selection presents some dynamical performance. An action selection model is established based on artificial life and MAS. The interaction and adaptation on perception and cognition of Agent are analyzed from dynamic system perspective; the action selection is regarded as the process to balance between adapting external environment and satisfying the internal desire. The simulation demonstrates the collective behavior of 21 Agents through local perception and interaction, and the result exhibits the global equilibrium and stability of collective behavior.

【Key words】perception; cognition; action selection; dynamic analysis

MAS(Multi-Agent System)主要研究Agent的组织结构和一定组织结构下的决策^[1], 能够描述Agent与环境和Agent之间的交互, 可以将其视为连接低级与高级或个体与群体行为的桥梁。将生命体视为Agent系统, 行为的产生被定义为行为选择或决策。Domenico^[2]指出, 人工智能将认知过程看作是对符号的处理, 没有考虑大脑、身体、环境等的物理机制, 而这些物理机制作为认知的基础不可忽略。人工生命则将认知理解为存在于某种物理环境下的生物现象, 运用具体的物理量建模, 更加符合生物体认知活动的自然本质。结合人工生命方法论和MAS, 运用其bottom-up原则, 是理解Agent行为选择动态性和揭示Agent低级反应与高级认知行为、简单个体与复杂群体行为内在关系的有效途径。

1 行为选择动态性分析

将个体 Agent 行为选择的基本过程描述为: 感知动态环境, 建立外部环境的感知评价函数; 根据内部知识, 建立内部意图的认知评价函数; 平衡动态的环境信息和内部知识, 进行行为选择。行为选择还受到 Agent 之间交互作用的影响, 文献[1]总结出“行为和交互是理解生命体并对其建模的主要工具”。交互可理解为两层含义: (1)Agent 与环境的信息交互和 Agent 之间的信息交互, 信息交互的结果使 Agent 获得超出自身感知能力和范围的外部信息; (2)Agent 之间交换各自的内部需求、意图或者评价准则。

行为选择则是一个动态平衡过程。适应外部环境和满足内部需求是 Agent 行为选择的目标, 属于一个多目标问题求解。可以视这一动态平衡过程为 Agent 的优化控制, 使系统趋于平衡和稳定。

2 个体 Agent 行为模型

根据上述动态特性, 受文献[3]的启发, 本文建立动态方

程, 用动态系统观点研究行为选择动态性和系统稳定性。

2.1 行为模型

设 Agent 在某一时刻有 N 种可选的行为 $i=1, 2, \dots, N$; 用 Agent 行为选择的可能性来表示行为状态空间, 设第 τ 步时, 选择行为 i 的可能性为 $x_i(\tau)$; 则第 τ 步 Agent 的行为状态向量表示为: $\mathbf{x}(\tau) = (x_1(\tau), x_2(\tau), \dots, x_N(\tau))$, 设置约束条件 $\sum_{i=1}^N x_i(\tau) = 1$ 。

设 Agent 在 τ 时刻选择行为 i 的收益为 $R_i(\tau)$, 所有行为的收益向量表示为

$$\mathbf{R}(\tau) = R_1(\tau), R_2(\tau), \dots, R_N(\tau)$$

设 Agent 在 τ 时刻根据感知到的外部环境信息, 对行为 i 的评价为 $P_i(\tau)$, 所有行为的感知评价向量为

$$\mathbf{P}(\tau) = P_1(\tau), P_2(\tau), \dots, P_N(\tau)$$

设 Agent 在 τ 时刻内部知识的认知结果, 对行为 i 的评价为 $C_i(\tau)$, 所有行为的认知评价向量为

$$\mathbf{C}(\tau) = C_1(\tau), C_2(\tau), \dots, C_N(\tau)$$

设 Agent 根据行为的收益来选择行为, 选择行为 i 的可能性为

$$x_i(\tau) = \frac{e^{\alpha R_i(\tau)}}{\sum_{n=1}^N e^{\alpha R_n(\tau)}} \quad (1)$$

其中, Agent 行为收益的改变受感知评价函数和认知评价函

基金项目:国家自然科学基金资助项目(60374032, 60503024)

作者简介:江道平(1978 -), 女, 博士研究生, 主研方向: 人工智能与人工生命; 班晓娟, 副教授、博士; 尹怡欣, 教授、博士、博士生导师

收稿日期:2007-04-15 **E-mail:** jdp_420@163.com

数的影响

$$R_i(\tau) - R_i(\tau-1) = \frac{1}{T} [\delta_i(\tau-1)P_i(\tau-1) + \sigma_i(\tau-1)C_i(\tau-1)] \quad (2)$$

其中, $\delta_i(\tau), \sigma_i(\tau)$ 分别控制感知评价和认知评价对收益的影响; α 控制行为受收益准则作用的适应度, 当 $\alpha \rightarrow \infty$ 时, Agent 拥有最大的 $R_i(\tau)$, 当 $\alpha \rightarrow 0$ 时, $x_i = 1/N$, 这时行为的选择与 $R(\tau)$ 无关, 表现为一种随机选择。

对于式(1)和式(2)的离散过程, 假设 Agent 进行感知评价和认知评价的时间远小于收益影响 Agent 行为并最终适应和稳定的过程, 可以将上述离散过程转换为连续过程。

由式(1)得

$$x_i(t) = \frac{e^{\alpha R_i(t)}}{\sum_{n=1}^N e^{\alpha R_n(t)}} \quad (3)$$

由式(2)得

$$\dot{R}_i(t) = \delta_i P_i(t) + \sigma_i C_i(t) \quad (4)$$

2.2 动态性分析

Agent 行为选择的目的是为了适应动态环境或者满足内部需求, 因此, Agent 的行为选择可以理解为适应环境和满足需求的一个动态平衡过程。

当 Agent 行为选择为适应外部环境, 即 Agent 的收益取决于外部环境信息, 则 $R_i(t) = P_i(t)$; 当 Agent 行为选择为满足内部需求, 即 Agent 的收益取决于内部知识, 则 $R_i(t) = C_i(t)$ 。

由式(3)和式(4)可得微分方程:

$$\frac{\dot{x}}{x} = \omega [P_i - \sum_{n=1}^N P_n x_n] + \nu [C_i - \sum_{n=1}^N C_n x_n] \quad (5)$$

该微分方程由 $P_i(t) - \sum_{n=1}^N P_n(t)x_n(t), C_i(t) - \sum_{n=1}^N C_n(t)x_n(t)$ 组成,

它们分别表示第 i 项行为的感知(认知)评价与所有行为的平均感知(认知)评价的差; ω, ν 分别控制感知、认知评价的权重, 通过调节系数可以控制 Agent 外部信息和内部知识的平衡, 使系统趋于稳定。

3 Agent 的交互

Agent 之间的交互使得外部环境信息更新、内部知识进化, 导致行为的修改, 对个体而言, 导致高级行为的涌现; 对于多 Agent 而言, 导致群体行为的涌现, 使系统表现出更强的动态性。

3.1 感知评价交互

设一 Agent x 的行为分布为 X , 收益函数为

$$\dot{R}_i^x(t) = \delta_i^x P_i^x(t) + \sigma_i^x C_i^x(t) \quad (6)$$

行为分布为

$$x_i(t) = \frac{e^{\alpha^x R_i^x(t)}}{\sum_{n=1}^N e^{\alpha^x R_n^x(t)}} \quad (7)$$

$$\dot{x}_i(t) = \alpha^x x_i(t) (\dot{R}_i^x(t) - \sum_{n=1}^N \dot{R}_n^x(t)x_n(t)) \quad (8)$$

$$\frac{\dot{x}_i}{x_i} = \omega^x [P_i^x - \sum_{n=1}^N P_n^x x_n] + \nu^x [C_i^x - \sum_{n=1}^N C_n^x x_n] \quad (9)$$

设另一 Agent 的行为分布为 Y , 收益函数为

$$\dot{R}_j^y(t) = \delta_j^y P_j^y(t) + \sigma_j^y C_j^y(t) \quad (10)$$

行为分布为

$$y_j(t) = \frac{e^{\alpha^y R_j^y(t)}}{\sum_{m=1}^M e^{\alpha^y R_m^y(t)}} \quad (11)$$

$$\dot{y}_j(t) = \alpha^y y_j(t) (\dot{R}_j^y(t) - \sum_{m=1}^M \dot{R}_m^y(t)y_m(t)) \quad (12)$$

$$\frac{\dot{y}_j}{y_j} = \omega^y [P_j^y - \sum_{m=1}^M P_m^y y_m] + \nu^y [C_j^y - \sum_{m=1}^M C_m^y y_m] \quad (13)$$

由于 Agent 交互的时间远小于 Agent 进行感知评价的时间, 可以认为在交互过程中 Agent 感知评价函数为常量, 即

$$P_{ij}^x = a_{ij}, P_{ji}^y = b_{ji}。$$

$$P_i^x = \sum_{m=1}^M a_{im} y_m, P_j^y = \sum_{n=1}^N b_{jn} x_n, \sum_{n=1}^N P_n^x x_n = \sum_{n=1}^N \sum_{m=1}^M a_{nm} x_n y_m, \sum_{m=1}^M P_m^y y_m = \sum_{n=1}^N \sum_{m=1}^M b_{mn} x_n y_m$$

经过感知评价交互以后, x, y 的行为动态方程表示为

$$\frac{\dot{x}_i}{x_i} = \omega^x [\sum_{m=1}^M a_{im} y_m - \sum_{n=1}^N \sum_{m=1}^M a_{nm} x_n y_m] + \nu^x [C_i^x - \sum_{n=1}^N C_n^x x_n] \quad (14)$$

$$\frac{\dot{y}_j}{y_j} = \omega^y [\sum_{n=1}^N b_{jn} x_n - \sum_{n=1}^N \sum_{m=1}^M b_{nm} x_n y_m] + \nu^y [C_j^y - \sum_{n=1}^M C_n^y y_n] \quad (15)$$

感知评价的交互, 使得 Agent $x(y)$ 获取的外部信息中包含自身感知信息和 Agent $y(x)$ 的行为信息, 更新了 $x(y)$ 的外部环境信息, 将 $y(x)$ 作为其外部环境的一个组成部分。

3.2 认知评价交互

由于 Agent 交互的时间远小于 Agent 进行认知评价的时间, 可以认为在交互过程中 Agent 认知评价函数为常量 $C_{ij}^x = c_{ij}, C_{ij}^y = d_{ij}$ 。

$$\text{则 } C_i^x = \sum_{m=1}^M c_{im} y_m, C_j^y = \sum_{n=1}^N d_{jn} x_n$$

$$\sum_{n=1}^N C_n^x x_n = \sum_{n=1}^N \sum_{m=1}^M c_{nm} x_n y_m, \sum_{m=1}^M C_m^y y_m = \sum_{n=1}^N \sum_{m=1}^M d_{nm} x_n y_m$$

$$\frac{\dot{x}_i}{x_i} = \omega^x [P_i^x - \sum_{n=1}^N P_n^x x_n] + \nu^x [\sum_{m=1}^M c_{im} y_m - \sum_{n=1}^N \sum_{m=1}^M c_{nm} x_n y_m] \quad (16)$$

$$\frac{\dot{y}_j}{y_j} = \omega^y [P_j^y - \sum_{m=1}^M P_m^y y_m] + \nu^y [\sum_{n=1}^N d_{jn} x_n - \sum_{n=1}^N \sum_{m=1}^M d_{nm} x_n y_m] \quad (17)$$

认知评价的交互, 使得 Agent $x(y)$ 的内部认知评价结果中包含自身的内部意图、准则和 Agent $y(x)$ 的行为信息, 更新了 $x(y)$ 的内部知识, 将 $y(x)$ 的内部需求作为其内部知识的一部分。

3.3 感知与认知交互

设 A, C 分别为 $N \times M$ 矩阵, B, D 分别为 $M \times N$ 的矩阵, 其中 $(A)_{ij} = a_{ij}, (B)_{ji} = b_{ji}, (C)_{ij} = c_{ij}, (D)_{ji} = d_{ji}$ 。如果同时进行感知评价交互和认知评价交互, 动态方程为:

$$\frac{\dot{x}_i}{x_i} = \omega^x [(Ay)_i - x \cdot Ay] + \nu^x [(Cy)_i - x \cdot Cy] \quad (18)$$

$$\frac{\dot{y}_j}{y_j} = \omega^y [(Bx)_j - y \cdot Bx] + \nu^y [(Dx)_j - y \cdot Dx] \quad (19)$$

将动态方程转换到另一个状态空间, 设 $\xi = \lg x, \zeta_i = \lg x_i,$

$\eta = \lg y, \eta_j = \lg y_j$ 。

$$\dot{\xi}_i = \omega^x (A_i e^{\eta} - e^{\xi} A e^{\eta}) + \nu^x (C_i e^{\eta} - e^{\xi} C e^{\eta}) \quad (20)$$

$$\dot{\eta}_j = \omega^y (B_j e^{\xi_j} - e^{\eta} B e^{\xi}) + \nu^y (D_j e^{\xi_j} - e^{\eta} D e^{\xi}) \quad (21)$$

交互后 Agent $x(y)$ 的环境信息是 $x(y)$ 环境感知和 $y(x)$ 行为的集合, 拓宽了 $x(y)$ 的感知范围和能力; 交互后 Agent $x(y)$ 的内部知识是 $x(y)$ 内部意图、准则和 $y(x)$ 行为的集合, $x(y)$ 的内部需求考虑了 $y(x)$ 传递的内部意图和准则。交互后 $x(y)$ 的行为选择是外部环境更新和内部知识进化之后, 适应外部环境和满足内部需求的平衡过程。由式(20)和式(21), 根据控制

原理，系统稳定点为 $\lim_{t \rightarrow \infty} e^{\xi(t)} = 0$ 或 $\lim_{t \rightarrow \infty} e^{\eta(t)} = 0$ ，可以由此来设定使系统达到平衡和稳定的系统参数。

4 动态性分析

4.1 没有交互

设 $P_i = a_i, C_i = c_i, N = M = 2, \omega = \nu = 1, \frac{\dot{x}_1}{x_1} = a_1 - (a_1 x_1 + a_2 x_2) + (a_1 x_1 + a_2 x_2) + c_1 - (c_1 x_1 + c_2 x_2)$

如果 $x_1 = x_2 = \frac{1}{2}$ 则 $\frac{\dot{x}_1}{x_1} = \frac{a_1 - a_2}{2} + \frac{c_1 - c_2}{2}$

$x_1 = 1, x_2 = 0$ 则 $\frac{\dot{x}_1}{x_1} = 0$

$x_1 = 0, x_2 = 1$ 则 $\frac{\dot{x}_1}{x_1} = (a_1 - a_2) + (c_1 - c_2)$

当 $a_1 = c_1, a_2 = c_2, \frac{\dot{x}_1}{x_1} = 0$

即 Agent 同等对待外部环境信息和内部知识时，行为选择表现出最小的动态性。

4.2 交互

假设 Agent x, y 同时进行感知评价交互和认知评价交互，并且 $N=M=2, \omega^x = \omega^y = \nu^x = \nu^y = 1, \frac{\dot{x}}{x} = \omega^x [(Ay)_i - x \cdot Ay] + \nu^x [(Cy)_i - x \cdot Cy]$ 。

$$\frac{\dot{x}_1}{x_1} = (A_{11}A_{12}) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - (x_1 x_2) \begin{pmatrix} A_{11}A_{12} \\ A_{21}A_{22} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + (C_{11}C_{12}) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - (x_1 x_2) \begin{pmatrix} C_{11}C_{12} \\ C_{21}C_{22} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$\frac{\dot{x}_1}{x_1} = ((A+C)_{11}(A+C)_{12}) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} - (x_1 x_2) \begin{pmatrix} (A+C)_{11}(A+C)_{12} \\ (A+C)_{21}(A+C)_{22} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

设 $A+C=E$

$$\frac{\dot{x}_1}{x_1} = E_{11}y_1 + E_{12}(1-y_1) - E_{11}x_1y_1 - E_{12}x_1(1-y_1) - E_{21}(1-x_1)y_1 - E_{22}(1-x_1)(1-y_1)$$

$$\text{当 } y_1 = y_2 = \frac{1}{2}, \frac{\dot{x}_1}{x_1} = \frac{(1-x_1)}{2} (E_{11} + E_{12} - E_{21} - E_{22})$$

$$x_1 = x_2 = \frac{1}{2}, \frac{\dot{x}_1}{x_1} = \frac{1}{4} [E_{11} + E_{12} - E_{21} - E_{22}]$$

$$x_1 = 1, x_2 = 0, \frac{\dot{x}_1}{x_1} = 0$$

$$x_1 = 0, x_2 = 1, \frac{\dot{x}_1}{x_1} = \frac{1}{2} [E_{11} + E_{12} - E_{21} - E_{22}]$$

$$y_1 = 1, y_2 = 0, \frac{\dot{x}_1}{x_1} = (1-x_1)(E_{11} - E_{12})$$

$$y_1 = 0, y_2 = 1, \frac{\dot{x}_1}{x_1} = (1-x_1)(E_{12} - E_{22})$$

$$\text{当 } E_{11} = E_{12} = E_{21} = E_{22}, \frac{\dot{x}_1}{x_1} = \frac{\dot{y}_1}{y_1} = 0$$

可见在 Agent 在同对待外部环境和内部知识的评价结果和交互 Agent 的相互影响时，行为选择表现出最小的动态性。

5 试验分析

5.1 试验仿真

通过 21 个 Agent 的两两交互，实现 Agent 群体和谐运动，即保持 Agent 以集群方式按相同速度运动，并防止碰撞。为

Agent 设置感知范围和交互范围，允许 Agent 感知自身感知范围内的环境信息和交互范围内的 Agent 进行两两交互。交互过程：(1)当位置坐标发生变化后，Agent i 根据交互范围更新交互成员库；(2)Agent i 同时发出信息给交互范围内的 Agent 并接受其他 Agent 发出的信息。

根据R-A(Repulsive-Attractive)模型^[6]，为Agent之间建立约束满足的搜索算法，采用break-out算法思想^[7]，使其跳出局部最小。

根据对外部感知评价和内部认知评价的平衡，运用通用行为选择模型从可行行为集中同步确定 Agent 的行为。算法求解具体过程见文献[8]。针对外部环境信息，当 Agent 违背权重较高的物理约束时，调高 ω ，使行为更多的适应动态环境；当出现权重较高的内部目标时，调高 ν ，使行为更多地满足内部需求，以维持系统平衡和稳定。

5.2 运行结果

运行结果如图 1~图 3 所示，表明 Agent 进行局部感知和局部交互，能够达到全局稳定。

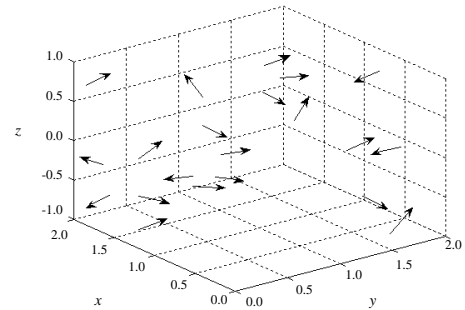


图 1 初始状态

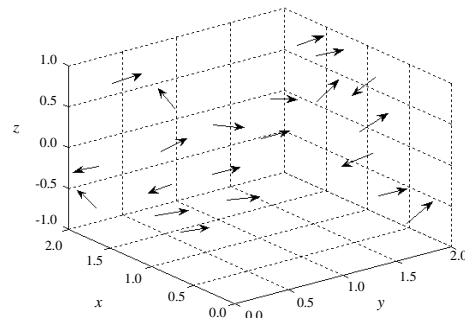


图 2 运行 100 步

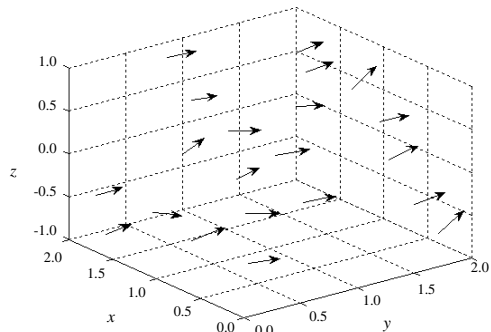


图 3 运行 500 步

(下转第 21 页)