

DOI: 10.13973/j.cnki.robot.2017.0680

基于自组织算法的情感机器人追捕任务分配

孙博寒, 王 浩, 方宝富, 凌兆龙, 林杰华

(合肥工业大学计算机与信息学院, 安徽 合肥 230009)

摘要: 提出了基于自组织映射 (self organizing map, SOM) 神经网络的自组织算法, 把任务分配融入网络训练过程中. 通过竞争获胜函数值在网络训练 (任务分配) 中的决定作用, 并结合算法自身任务分配的实时性, 由各情感机器人根据情感等因素提供竞争获胜函数值, 并对值进行强化学习调整. 这样使情感直接有效地参与任务分配决策, 优化了算法性能. 最后, 通过仿真实验验证了本文所提出算法的有效性, 特别是随着情感机器人团队规模扩大, 追捕时间会比现有算法缩短一半以上.

关键词: 自组织映射神经网络; 情感机器人; 任务分配; 强化学习; 竞争获胜函数

中图分类号: TP242

文献标识码: A

文章编号: 1002-0446(2017)-05-0680-08

Task Allocation in Emotional Robot Pursuit Based on Self-organizing Algorithm

SUN Bohan, WANG Hao, FANG Baofu, LING Zhaolong, LIN Jiehua

(School of Computer and Information, HeFei University of Technology, Hefei 230009, China)

Abstract: A self-organizing algorithm based on self-organizing map (SOM) neural network is proposed and task allocation is integrated into the training process of the network. Taking advantages of the decisive effect of the competitive winning function value in network training (task allocation) and combining with the real-time performance of task allocation in the proposed algorithm, the emotional robot provides the competitive victory function value, which is adjusted by reinforcement-learning, according to the emotional factors. In this way, the emotion directly and effectively participates in task allocation decisions, and the algorithm performance is optimized. Finally, simulation experiments are carried out to verify the effectiveness of the proposed algorithm. Especially, with the scale of emotional robot team growing, the pursuit time can be shortened by more than 50% compared with the existing algorithm.

Keywords: SOM (self-organizing map) neural network; emotional robot; task allocation; reinforcement learning; competitive winning function

1 引言 (Introduction)

近年来, 随着人工情感理论研究的开展与机器人相关技术的成熟, 情感机器人应运而生. 情感机器人的研究, 很多是在人机交互方面, 如人面部表情、手势的识别表达^[1]. 而有些学者把情感引入到多机器人协作中, 发挥了情感的决策作用, 提升了多机器人系统的性能, 如多智能体的控制方面^[2]、多机器人追捕任务分配^[3]、多智能体决策系统^[4]. 这些学者为了把情感用于多机器人协作, 提出了不同的情感认知模型, 包括: (1) 大五人格理论 (OCEAN) 模型^[5], 把个性划分为 5 个维度, 开放性、尽责性、外向性、宜人性和神经质. (2) 情感协作因子模型^[3], 对状态空间与概率空间极大相似度匹配模型进行扩充, 用情感协作因子量化协作意

愿.

任务分配是多机器人协作系统的研究热点之一, 多机器人任务分配的目的在于确保所有任务可以完成的基础上, 把不同的子任务分配给机器人团队中各成员, 使得系统效率及完成任务收益最大化. 特别是, 国内外有些学者对情感机器人的任务分配进行了相关研究. 文 [3] 研究了基于情感协作因子的多机器人追捕任务分配, 提出了情感协作因子模型, 利用情感协作因子衡量机器人的协作意愿, 其大于一定阈值则使用拍卖算法进行任务分配. 文 [6] 研究了使用复合情感算法的多机器人任务分配^[6], 为每个机器人建立“生气”和“感恩”2 个情感基因, 情绪受这 2 个情感基因影响, 达到一定阈值则响应广播的任务. 文 [7] 对多机器人情

感智能系统建立合作计划进行任务分配, 每个机器人建立 4 个基本情感 (高兴, 生气, 恐惧, 悲伤) 状态, 采用马尔可夫情感模型决定情感变化, 机器人根据恐惧值决定是否寻求帮助, 根据悲伤值决定是否参与合作. 这些算法在处理情感认知模型与任务分配关系时, 考虑不完善, 或者不适合处理大规模情感机器人追捕团队的任务分配.

本文在文 [3] 的情感认知模型基础上, 为情感机器人建立基于个性的情感状态转移模型, 并通过 SOM 神经网络^[8] 及其训练过程进行任务分配. 情感机器人结合自身环境因素 (位置、收益等) 与情感因素产生竞争获胜函数值, 通过 SOM 神经网络^[8] 训练, 使得情感直接参与到任务分配决策中. 对竞争获胜函数值进行强化学习调整, 可优化分配性能. 这样, 既充分发挥了情感的决策作用, 以及情感机器人的自主性, 又提升了在规模较大的情感机器人追捕团队中进行任务分配的性能.

2 情感机器人追捕问题任务场景 (Task scenarios of the emotional robot pursuit problem)

追捕问题是体现多情感机器人系统合作完成任务的一个典型问题, 非常适合任务分配的研究. 在现实环境中, 例如警察追捕逃犯时, 热心尽责的警察可能会先帮助队友围堵逃犯, 然后再追捕其他逃犯, 这样有利于集中警力稳步地完成任任务. 基于此, 将情感赋予机器人进行追捕问题研究, 使其有人在决策过程中体现的自主性、灵活性.

2.1 机器人描述

定义 1 (逃跑机器人) 作为追捕机器人追捕的任务, 定义为一个三元组 $E_i = \langle L_i, C_{Ei}, R_i \rangle$, 其中 L_i 代表逃跑者在追捕场景中的位置, 为 2 维平面坐标 (x, y) ; C_{Ei} 代表逃跑机器人的能力值, 与追捕难度成正比; R_i 代表此逃跑机器人所具有的价值, 是捕获被追捕机器人后得到的报酬 (可以通过此值设定优先级).

定义 2 (追捕机器人) 具有情感的机器人, 定义为五元组 $P_j = \langle L_j, C_{Pj}, Q_{ej}, Q_{pj}, B_j \rangle$, 其中: L_j 代表追捕者在追捕场景中的位置, 为平面坐标 (x, y) ; C_{Pj} 代表追捕机器人的能力值, 与完成任务能力成正比; Q_{ej} 代表追捕机器人当前的情感状态, 表示为 $Q_{ej} = [e_1, e_2, \dots, e_N]$, 其中 e_i 表示某一基本情感^[5], $e_i \in [0, 1]$; Q_{pj} 代表追捕机器人的个性值, 表示为 $[O, C, E, A, M]$ ^[9]; B_j 代表机器人的情感决策部分, 提供竞争获胜函数值.

2.2 任务场景

本文的追捕任务场景设定在一个无界的 2 维平面, 由 M 个追捕机器人追捕 N 个逃跑机器人, 2 种类型的机器人位置随机选定. 在追逃过程中, 机器人不限定方向, 只设定运动的步长 (速度).

定义 3 当 $C_{si} = 1$ 时判定逃跑机器人 i 被捕获, $C_{si} = 0$ 时判定为捕获失败, 其中:

$$C_{si} = \begin{cases} 1, & \forall D_{Pj} \leq r, \sum_j C_{Pj} \geq C_{Ei} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

式中 r 为捕获范围, D_{Pj} 为追捕机器人 j 与目标逃跑机器人之间的距离. 如图 1 所示, 当追捕者 P_1 、 P_2 、 P_3 移动到逃跑者 E_1 的半径为 r 的圆域内时, $\sum_{j=1}^3 C_{Pj} \geq C_{E1}$, 逃跑者 E_1 被捕获.

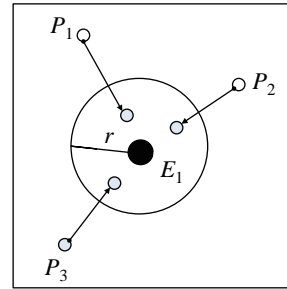


图 1 追捕示意图

Fig.1 Schematic diagram of pursuit

逃跑机器人被成功抓住后, 追捕机器人则被释放出来继续追捕其他逃跑机器人, 当所有逃跑机器人都被成功抓住后, 整个追捕任务完成.

定义 4 团队追捕完成时间 T_{total} 为捕获所有逃跑机器人的总时间, 公式如下:

$$T_{total} = \sum_{i=1}^N (\max(T_{end}) - T_{start})_i \quad (2)$$

$$T_{min} = \min T_{total} \quad (3)$$

其中 T_{start} 为逃跑者 i 开始逃跑的时间, T_{end} 为逃跑者 i 被抓住的时间. T_{min} 是算法优化的目标函数, 代表完成所有任务所需的最短时间. 本文把 T_{min} 作为评价追捕任务分配算法的标准.

3 情感模型及情感决策过程 (Emotional model and the process of emotional decision)

因为机器人赋予情感后有自主与灵活特性, 为了发挥情感决策作用, 本文对追捕机器人建立基于个性的情感模型.

在 SOM 神经网络的训练过程中，竞争获胜函数值决定每一轮获胜神经元，最终影响结果，见文 [8]。本文基于 SOM 神经网络^[8] 及训练过程进行任务分配，竞争获胜函数值会影响任务分配结果。情感作为生成竞争获胜函数值的因素之一，通过影响竞争获胜函数值的方式，可以直接参与到任务分配决策中。

3.1 基于个性的机器人情感模型

本文的情感更新模型主要包括情感刺激模型和情感衰减模型，在任务分配过程中，情感会不断更新，进而影响情感协作因子^[3] 的变化。其中还考虑了个性的因素。完整的机器人情感模型如图 2 所示。

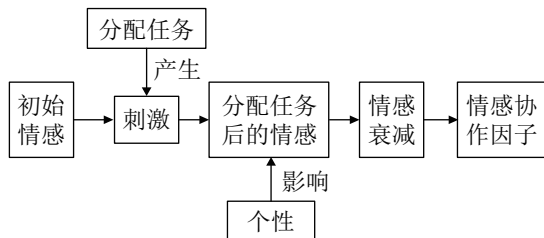


图 2 基于个性的情感模型

Fig.2 Emotional model based on personality

以下是 2 个组成部分介绍：

(1) 情感衰减模型方面，心理学中所提到的情感强度第 3 定律指出，情感衰减所对应的曲线与指数函数 $y = e^{-x}$ 最为吻合。为简化情感衰减过程，设定此时的衰减情况只与上一时刻有关，定义如下公式：

$$\phi(k_i) = e^{-k_i T} \tag{4}$$

$$\mathbf{Q}_{e_t} = \mathbf{Q}_{e_{(t-1)}} \times \phi(k_i) \tag{5}$$

$\mathbf{Q}_{e_{(t-1)}}$ 为 $t-1$ 时刻的情感向量， T 为设定的衰减周期， k_i 为机器人的个性参数，不同个性有不同的衰减情况。

(2) 情感刺激模型方面，定义了刺激矩阵^[10]，用刺激作为促进情感转化的因素，定义 I_T 表示刺激的集合，刺激的种类应该与情感 $\mathbf{Q}_{e_t} = [e_1, e_2, \dots, e_N]$ 中基本情感 e_i 相对应，即 $I_T = \{I_{s1}, I_{s2}, \dots, I_{sN}\}$ ，其中 I_{si} 表示对某一基本情感 e_i 的刺激量，数值越大表示刺激越强烈且 $I_{si} \in [-1, 1]$ 。本文只考虑捕获逃跑者所获的报酬对追捕者的情感刺激。同一个逃跑者无论被哪个追捕者成功捕获，其报酬都相同，表示为 R_j 。并且，把所有逃跑者报酬组成的序列 $(R_1, \dots, R_j, \dots, R_N)$ 归一化，这样产生出报酬的刺激序列 $(I_{T1}, \dots, I_{Tj}, \dots, I_{TN})$ ，其中 $I_{Tj} = R_j \times$

$(I_{s1}, I_{s2}, \dots, I_{sN})$ 。结合影响极限概率 π^* 的追捕者 i 的个性因素，经过 HMM (隐马尔可夫模型) 随机过程后^[11]，可以得到追捕者 i 分配到逃跑者 j 时由任务报酬产生刺激后的情感状态。最后，计算追捕者 i 的情感协作因子^[3]，作为追捕者 i 追捕逃跑者 j 的意愿量化值。

3.2 情感决策过程

本文中，情感决策过程作为定义 2 中追捕机器人的 B_j 部分，把情感作为竞争获胜函数值生成因素之一，影响任务分配结果。

追捕机器人 m 的情感决策过程如下：

Step1 建立如图 3 所示的 3 层 BP (反向传播) 神经网络^[12]，输入层神经元个数为 $2 \times N$ ， N 为逃跑者数量。输入值可以表示成 $\mathbf{s}_i = (d_1, f_1, \dots, d_n, f_n, \dots, d_N, f_N)$ 形式，其中 d_n 表示该追捕者到逃跑者 n 的距离， f_n 表示该追捕者对于逃跑者 n 的情感协作因子^[3]。输出层神经元的个数为 N ， N 为逃跑者数量。设定隐层节点数为 $H = I + L + \sigma$ ，其中 I 和 L 是输入和输出层的神经元数量， σ 为 1~10 之间的常数。激活函数选取的是 S 型函数： $f(x) = \frac{1}{1 + e^{-\alpha x}}$ ($0 \leq f(x) \leq 1$)， α 为 0.5。

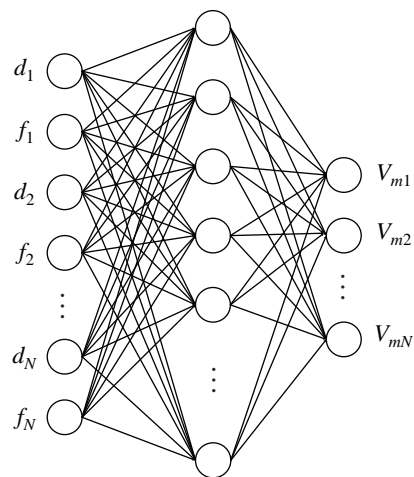


图 3 3 层 BP 神经网络结构

Fig.3 The structure of three-layer BP neural network

Step2 追捕者 m 从 SOM 样本训练中收到的任务值 i ，根据自身环境与情感状态，产生情感协作因子 f_n 与任务距离 d_n 。组成 $\mathbf{s}_i = (0, 0, \dots, d_n, f_n, \dots, 0, 0)$ 作为每个追捕者 BP 神经网络的当前输入值，每个输入值也是 3.3 节中强化学习的一个状态值。

Step3 输出值为 $(V_{m1}, \dots, V_{mi}, \dots, V_{mN})$ ， V_{mi} 就是追捕者 m 对任务值 i 的竞争获胜函数值。每个追捕者把相应任务的竞争函数值反馈给 SOM 神经

网络^[8], 也代表追捕者结合自身情况做出的量化选择.

3.3 竞争获胜函数值调整

因为不知道情感因素、距离因素与竞争获胜函数值存在何种关系使任务分配结果最优, 所以通过强化学习调整竞争获胜函数值来优化结果. 情感因素、距离因素作为状态是连续的, 而表格式 Q 学习算法只能应用到离散状态中, 为此采用基于 BP 神经网络的 Q 学习算法来实现连续状态的强化学习^[12]. 这就是 3.2 节为每个追捕者建立 BP 神经网络的原因.

Q 强化学习^[12] 利用类似 Q 函数的形式进行学习更新, 采用以下公式更新 Q 值:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r(s_t, a_t) + \gamma \max_{a \in A} Q(s_{t+1}, a) - Q(s_t, a_t)] \quad (6)$$

本文中, 状态 s_t 为 BP 神经网络的输入值 s_t , $Q(s_t, i)$ 为 BP 神经网络的输出值 V_{mi} , $i \in A$ (A 为逃跑者集合). 首先, 每个追捕者初始化 BP 神经网络的权值和所用到的参数, 结合 BP 训练算法^[12] 调整权值逼近向量函数:

$$C(s_i) = (0.5 \cdot d_1 + 0.5 \cdot f_1, \dots, 0.5 \cdot d_n + 0.5 \cdot f_n, \dots, 0.5 \cdot d_N + 0.5 \cdot f_N) \quad (7)$$

其中, N 为逃跑者个数, 使 BP 神经网络具有初始竞争函数作用. 然后执行以下步骤完成整个强化学习调整过程:

Step1 设定 Q 强化学习中的 α 学习因子、 γ 折扣因子等参数.

Step2 追捕者 m 从 SOM 训练中收到任务值 i , 结合 3.2 节的情感决策过程, 获得当前的状态 s_t 与 $Q(s_t, i)$, $i \in A$ (A 为逃跑者集合).

Step3 若该追捕者被分配任务, 则记录 $Q(s_t, i)$ 作为当前状态/行为评价; 否则, 返回 **Step2**, 不进行调整.

Step4 在 SOM 的训练过程中, SOM 神经网络权值的调整使得该追捕者位置变化. 追捕者 m 从 SOM 训练中收到下一任务值, 再次结合 3.2 节的情感决策过程, 获得一个新的状态 s_{t+1} , 同时获得强化信号 (环境奖励值).

Step5 利用式 (6) 更新 Q 函数, 并使用 BP 算法^[12] 训练神经网络.

Step6 重复 **Step2** ~ **Step5** 直到完成追捕任务, 完成一次学习过程.

神经网络强化学习的调整由每个追捕者独立进行, 不占用整体的追捕时间代价.

4 追捕任务分配自组织算法 (Self-organizing algorithm of pursuit task allocation)

本文的自组织算法基于 SOM 神经网络, 解决了情感机器人的追捕问题. 把任务分配与追捕过程融入到 SOM 神经网络的训练过程中, 并转变其中竞争获胜函数值的产生方式. 竞争获胜神经元及邻域内所有神经元所代表的追捕者为任务分配结果, 追捕过程对应相关神经元的权值调整. 以下是算法的具体描述.

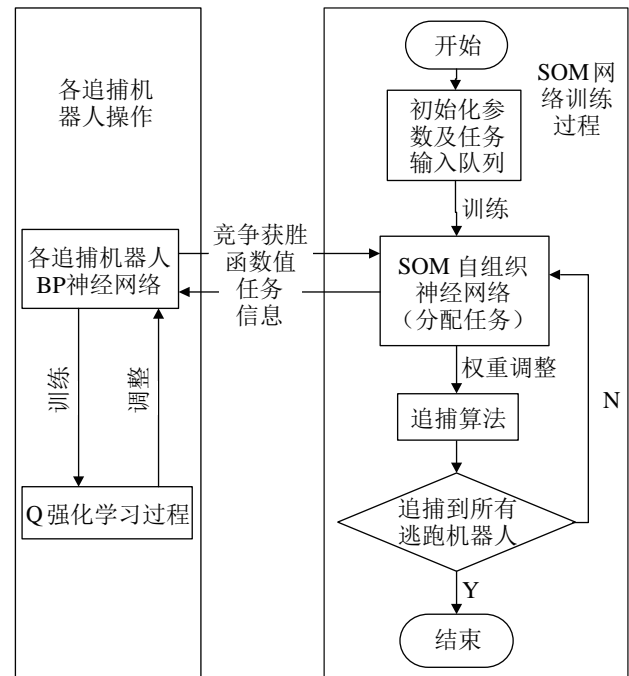


图 4 自组织算法流程图

Fig.4 Flow chart of the self-organizing algorithm

4.1 算法流程

本文提出的情感机器人任务分配自组织算法主体思想: 把逃跑者 (任务) 信息根据优先级排列, 循环输入 SOM 神经网络, 根据 SOM 神经网络的自组织算法训练, 完成任务分配与抓捕过程. 整体流程:

Step1 首先赋予追捕机器人情感、个性模型, 然后为每个追捕机器人建立各自的 BP 神经网络并且训练初始化. 设定自组织网络训练的输入任务队列 (逃跑者位置).

Step2 SOM 神经网络^[8] 把输入值分别发给各个追捕者. 追捕者根据不同的逃跑者 (任务) 报酬刺激, 会产生不同的情感协作因子, 并结合距离值

组合成状态输入到各自的 BP 神经网络，然后把结果反馈给 SOM 神经网络作为竞争获胜函数值。

Step3 当 SOM 神经网络收到所有追捕者的竞争函数值后，计算出获胜的神经元及邻域完成任务分配，然后调整邻域内神经元权值（追捕过程）。同时，每个追捕者根据任务分配结果对竞争获胜函数值进行强化学习调整

Step4 当抓捕到所有逃跑者时，SOM 训练过程结束；否则，转到 **Step2** 继续训练。整个任务分配过程如图 4 所示。

4.2 SOM 神经网络结构

建立 2 层 SOM 神经网络结构^[13-14]，如图 5 所示：(1) 输入层包括 K 个神经元 (x_1, x_2, \dots, x_K) ，代表逃跑者在追捕场景中的坐标 L_i ，本文设定为 2 维，即 $K = 2$ 。所有逃跑者的坐标构成输入集合，并且依次排列，排列顺序由逃跑者的优先级决定。(2) 竞争层（输出层）包括 $M \times N$ 个神经元，可表示为 $(R_{11}, R_{12}, \dots, R_{1N}, \dots, R_{m1}, R_{m2}, \dots, R_{mN}, \dots, R_{M1}, R_{M2}, \dots, R_{MN})$ ， N 为逃跑机器人的个数， M 为追捕机器人的个数。其中 N 个神经元为一组，第 m 组代表第 m 个追捕机器人的任务分配情况。输入层神经元与输出层神经元全连接，输入层到输出层神经元的连接 $x_k \rightarrow R_{mn}$ 的权值表示为 w_{kmn} ，将连接到输出层每个神经元的权值组 $W_{mn} = (w_{1mn}, w_{2mn}, \dots, w_{Kmn})$ 初始化为追捕机器人初始位置的坐标。

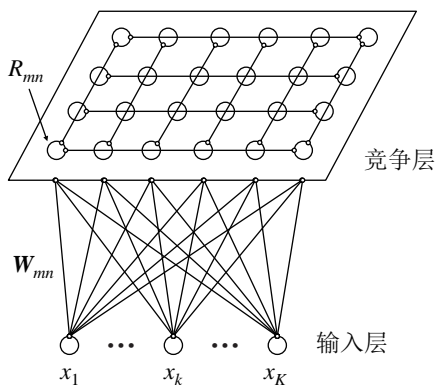


图 5 2 层 SOM 神经网络
Fig.5 The two-layer SOM neural network

4.3 神经元竞争获胜标准

用 SOM 竞争层神经元代表各追捕机器人的任务分配情况，获胜表示追捕者被分配任务。在 SOM 训练中，将神经元与输入模式向量的相似度作为竞争获胜的评判标准。本文 SOM 竞争层神经元权值向量用来提供与记录追捕者的位置，不直接用于相似度比较，而使用追捕者通过情感决策过程反馈的竞争获胜函数值 V_{mi} 来代表与输入模式向量的相似

度。

训练时输入模式向量为逃跑者的位置坐标 L_i ，定义 D_{mni} 表示神经元 R_{mn} 与输入 L_i 的相似值，公式如下：

$$D_{mni} = \frac{V_{mi}}{\partial_{mn}} \tag{8}$$

其中， V_{mi} 为追捕者 m 反馈的竞争获胜函数值， ∂_{mn} 为神经元 R_{mn} 的抑制参数，随着神经元权值被调整的次数增多而减小，保证每个机器人（神经元）都有机会参与到追捕任务中。根据 SOM 训练算法， $D_{m^*n^*i}$ 表示神经元 $R_{m^*n^*}$ 对输入 L_i 的相似值。当 $D_{m^*n^*i} = \max(D_{mni})$ ， $1 \leq m \leq M$ ， $1 \leq n \leq N$ 时，则竞争层神经元 $R_{m^*n^*}$ 获胜。

4.4 获胜邻域

由竞争获胜函数值得出获胜的输出层神经元后，就要对邻域内神经元作调整。这个过程就是组建追捕团队，所以获胜邻域就是设定追捕团队成员。根据之前的描述，追捕机器人团队的能力值大于逃跑者的能力值，这个团队才能成功抓捕逃跑者，所以邻域函数定义为

$$f(d_{mn}) = \begin{cases} e^{-d_{mn}^2}, & d_{mn} \leq r \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

其中： $0 \leq f(d_{mn}) \leq 1$ ； $d_{mn} = \|R_{m^*n^*} - R_{mn}\| = |D_{m^*n^*i} - D_{mni}|$ 为第 m 组中第 n 个输出神经元与获胜神经元 $R_{m^*n^*}$ 的距离； r 为邻域半径，若邻域内神经元（追捕者团队）能力值大于等于输入的逃跑者 i 的能力值，则邻域内最远神经元的距离就是邻域半径。

4.5 权值调整与追捕算法

在 SOM 神经网络中，确定获胜神经元和邻域内神经元后，就要进行权值调整。而对应于追捕问题，竞争层神经元的权值代表追捕者的坐标，权值调整这个过程相当于任务分配后的追捕机器人团队进行追捕。权值调整规则如下：

$$W_{mn}(t+1) = \begin{cases} W_{mn}(t), & \text{邻域外} \\ W_{mn}(t) + L_{\text{step}} \cdot f(d_{mn}) \cdot g(L_i, W_{mn}(t)), & \text{邻域内} \end{cases} \tag{10}$$

其中 L_{step} 为权值调整步数， $f(d_{mn})$ 为邻域函数， $g(L_i, W_{mn}(t))$ 表示为追捕者 m 运动的速度矢量（权值调整的矢量方向单位步长）。因为采用了追捕算法获得权值调整方向，最终一定能完成追捕任务，所以这样修改权值不影响 SOM 训练算法的收敛性。

追捕者速度矢量的设定使用了追捕算法, 本文实验应用了文 [3] 的人工势场法:

$$F_a = \gamma \frac{1}{d_{E_i}} \quad (11)$$

γ 为比例系数, d_{E_i} 为追捕机器人到逃跑者 E_i 的距离. 追捕者选取引力 F_a 最大的方向单位步长为速度矢量.

同时, 本文中逃跑者的逃跑原则, 也采用了人工势场法:

$$F_r = \gamma \left(\frac{1}{d_{P_1}} + \frac{1}{d_{P_2}} + \dots + \frac{1}{d_{P_n}} \right) \quad (12)$$

γ 为比例系数, d_{P_i} ($i = 1, 2, \dots, n$) 表示追捕团队中的机器人到所分配逃跑者的距离. 逃跑者选取斥力 F_r 最小的方向运动.

5 实验与结果 (Experiments and results)

在实验中, 随机场景设定为: 追捕者的速度略大于逃跑者, 追捕者数量大于逃跑者的数量. 逃跑者的位置及报酬属性随机设定, 追捕者的位置、情感及个性属性 (设置 3 维基本情感高兴、生气、恐惧的强度值作为情感值) 随机设定, 其能力值可使平均 3 个追捕者能抓住 1 个逃跑者.

式 (8) 中的 ∂_{mn} 抑制参数初值设定为 1, 神经元 R_{mn} 每获胜一次 ∂_{mn} 减 0.05. Q 强化学习中的 α 学习因子为 0.01、 γ 折扣因子为 0.95, 强化信号为:

追捕者没抓到逃跑者时信号为 r_{nc} :

$$r_{nc} = \begin{cases} -0.05, & \text{远离一个逃跑者} \\ +0.05, & \text{接近一个逃跑者} \\ 0, & \text{其他} \end{cases}$$

追捕者抓住逃跑者时信号为 r_{ce} :

$$r_{ce} = \begin{cases} +0.11, & \text{比上一次抓到相对应的逃跑者快} \\ -0.11, & \text{比上一次抓到相对应的逃跑者慢} \\ 0, & \text{其他} \end{cases}$$

以上参数设定是既考虑自组织算法的稳定性, 又充分发挥强化学习作用的方案. 以完成追捕的总时间 T_{total} 作为评价标准, 进行如下实验, 验证本文算法.

实验 1 验证基于自组织算法的情感机器人追捕任务分配的可行性, 特别是其中竞争函数强化学习的调整效果. 实验中, 设定一个 9 追 3 随机场景, 并且多次重复实验, 得出如图 6 的结果.

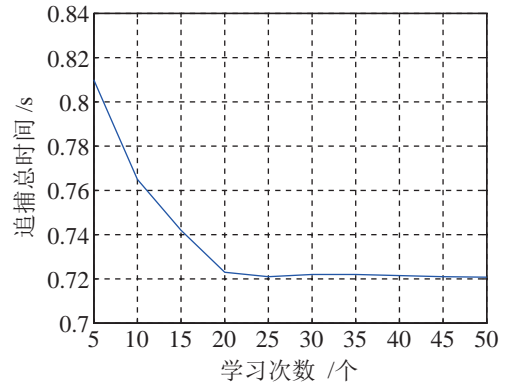


图 6 强化学习对算法性能的调整
Fig.6 Adjustment of the algorithm performance by reinforcement-learning

从图 6 可以看到, 追捕任务都顺利地完成了, 而且竞争函数强化学习对于追捕任务完成起到了调整的作用. 因为本文算法的自组织特性, 强化学习起到了优化作用, 但不能从本质上提升性能. 在该试验场景下训练次数为 20 次左右时, 算法性能趋于稳定.

实验 2 验证情感机器人任务分配算法比理性机器人任务分配更灵活与自主, 以及本文算法对情感机器人任务分配的优势. 设定一个 9 追 3 随机场景, 使用本文的情感机器人任务分配算法与理性机器人拍卖算法^[15] 进行抓捕, 结果如图 7 所示.

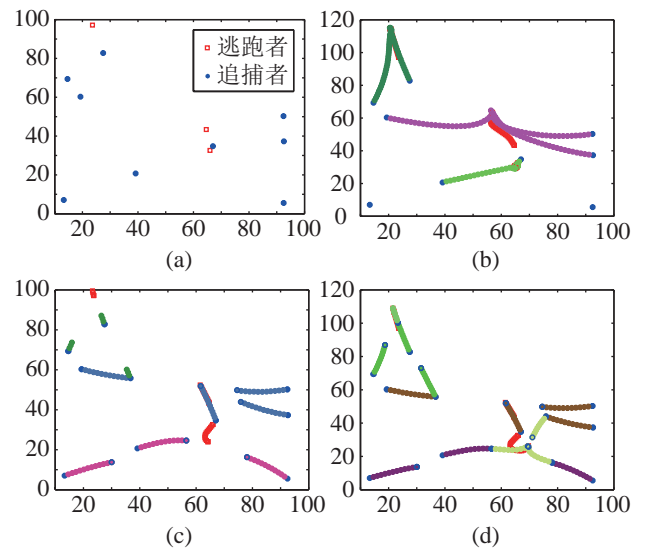


图 7 追捕过程对比图
Fig.7 The comparison chart of pursuit process

通过对实验结果的分析, 在图 7(a) 场景下: (1) 理性机器人任务分配如图 7(b) 所示, 追捕者 1、2 会一直追捕初始指派的最优目标, 直到任务完成. (2) 本文情感机器人任务分配, 追捕者受自身情感影响, 初始时不一定选择距离最近的目标, 如 7(c)

中追捕者 1 选择追捕逃跑者 2, 而不是最近的逃跑者 1. 而且追捕过程中, 追捕者情感变化与本文算法的分配实时性结合, 会影响分配的目标; 如追捕者 2 先追捕逃跑者 2 (图 7(c)), 后来分配追捕逃跑者 3 (图 7(d)), 因为追捕者 2 此时恰巧离逃跑者 3 较近, 反而加速了追捕过程, 比理性机器人先完成追捕任务.

实验 3 验证本文的自组织算法适合处理大规模情感机器人追捕任务分配的特点. 为不同规模的情感机器人追捕任务各设定 100 种随机场景. 同时运行本文的自组织算法和基于情感协作因子的多机器人追捕任务分配算法^[3] 处理不同规模的追捕任务, 结果如图 8 所示.

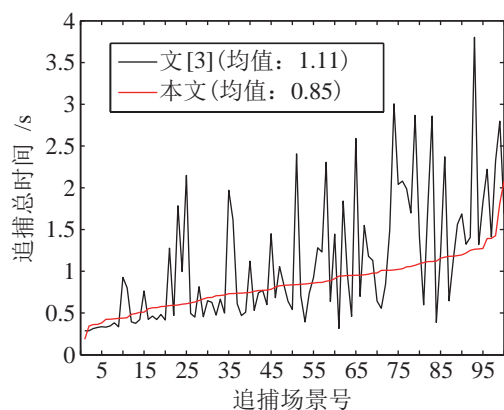


图 8 9 追 3 规模的追捕效率对比

Fig.8 The comparison of pursuit efficiency on 9-pursuit-3 scenarios

由图 8 可以看出, 在对 9 追 3 这种小规模情感机器人追捕任务进行任务分配时, 本文算法与文 [3] 算法在完成追捕的时间上差别不大, 性能相似.

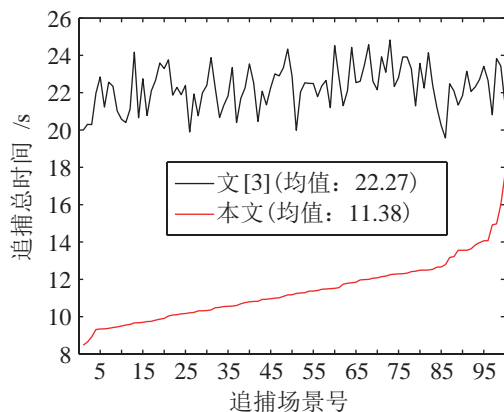


图 9 60 追 20 的追捕效率比

Fig.9 The comparison of pursuit efficiency on 60-pursuit-20 scenarios

由图 9 可以看出, 在对 60 追 20 这种较大规模

追捕任务进行任务分配时, 本文算法比文 [3] 算法完成追捕的时间要缩短一半以上, 体现了本文的自组织算法在对情感机器人大规模追捕场景进行任务分配时的优势.

由图 10 可以看出, 通过比较不同规模的情感机器人的 100 种随机场景的追捕时间均值, 本文算法随着追捕任务中情感机器人的规模增大, 追捕平均总时间明显优于文 [3] 中的算法. 而且从图 11 可以看出, 本文算法的追捕时间方差基本上低于文 [3] 中的算法, 说明本文算法对情感机器人任务分配的稳定性好于文 [3] 中的算法.

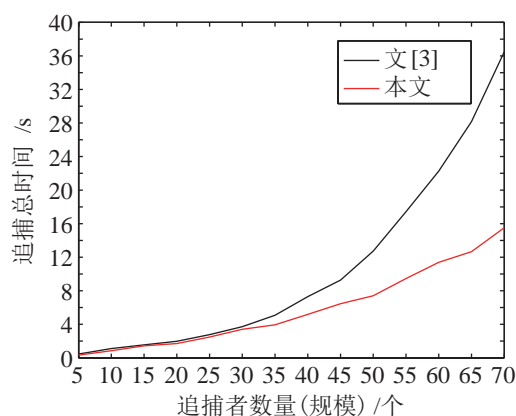


图 10 平均追捕效率比

Fig.10 The comparison of average pursuit efficiency

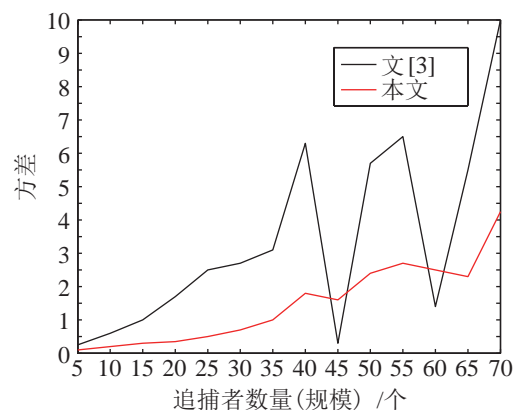


图 11 追捕时间方差比较

Fig.11 The comparison of pursuit time variance

具体分析实验结果后发现: (1) 本文把情感协作因子作为竞争获胜函数的输入参数之一, 相比于文 [3] 算法, 使情感更好地融入任务分配中. (2) 文 [3] 的拍卖算法中使用匈牙利算法获得全局最优分配. 本文的自组织算法具有局部最优选择特点, 降低了对大规模机器人团队任务分配的时间复杂度. (3) 文 [3] 中的算法只有在抓到逃跑者或者愿意协作的追捕者少时, 才进行重分配, 实时性较低.

本文的自组织算法任务分配与追捕算法交替进行, 使得任务分配具有实时性。

6 结论 (Conclusion)

本文基于 SOM 神经网络及其训练算法提出了一种情感机器人追捕任务分配的自组织算法。情感机器人由情感因素与环境因素 (位置、收益等) 得出竞争获胜函数值, 影响任务分配结果。各个追捕者独立并行地对竞争获胜函数值进行强化学习, 既优化了任务分配, 又不影响算法的整体效率。与其他情感机器人任务分配算法相比, 本文算法把情感融入了任务分配过程中, 结合任务分配的实时性特点, 更好地发挥了情感的作用。而且在处理大规模情感机器人任务分配时, 追捕时间比现有算法缩短一半以上。下一步将研究适合本文自组织算法的情感机器人追捕算法, 优化追捕过程, 进一步提升性能。

参考文献 (References)

- [1] Oh K G, Jang M S, Kim S J. Automatic emotional expression of a face robot by using a reactive behavior decision model[J]. *Journal of Mechanical Science and Technology*, 2010, 24(3): 769-774.
- [2] Murphy R R, Lisetti C L, Tardif R, et al. Emotion-based control of cooperating heterogeneous mobile robots[J]. *IEEE Transactions on Robotics and Automation*, 2002, 18(5): 744-757.
- [3] Fang B F, Chen L, Wang H, et al. Research on multirobot pursuit task allocation algorithm based on emotional cooperation factor[J]. *The Scientific World Journal*, 2014, 2014: No.864180.
- [4] Watada S, Obayashi M, Kurenoto T, et al. A new decision-making system of an agent based on emotional models in multi-agent system[C]//18th International Symposium on Artificial Life and Robotics. Oita, Japan: Alife Robotics Co, Ltd, 2013: 452-455.
- [5] 谷学静, 王志良, 魏哲华, 等. 基于人工心理理论的情感模型构建方法研究 [J]. *微计算机信息*, 2006, 22(5): 264-266.
Gu X J, Wang Z L, Wei Z H, et al. The research of constructing affective model based on artificial psychology theoretics[J]. *Microcomputer Information*, 2006, 22(5): 264-266.
- [6] Yuan W, Zeng B. Multi-robot task allocation using compound emotion algorithm[C]//7th International Symposium on Advanced Parallel Processing Technologies. Berlin, Germany: Springer-Verlag, 2007: 545-550.
- [7] Banik S C, Watanabe K, Izumi K. Task allocation with a cooperative plan for an emotionally intelligent system of multi-robots[C]//SICE Annual Conference. Piscataway, USA: IEEE, 2007: 1004-1010.
- [8] Kohonen T. Self-organizing maps: Optimization approaches[C]//1991 International Conference on Artificial Neural Networks. Amsterdam, Netherlands: Elsevier, 1991: 981-990.
- [9] Digman J M. Personality structure: Emergence of the 5-factor model[J]. *Annual Review of Psychology*, 1990, 41(1): 417-440.
- [10] Wang G J, Teng S D, Fu K C. Simulating emotion and personality for intelligent agent[C]//International Conference on Intelligent Computation Technology and Automation. Piscataway, USA: IEEE, 2010: 304-308.
- [11] 王浩, 张权益, 方宝富, 等. 基于状态空间与概率空间映射的极大相似度匹配情感模型 [J]. *模式识别与人工智能*, 2013, 26(6): 552-560.
Wang H, Zhang Q Y, Fang B F, et al. Maximum similarity matching emotion model based on mapping between state space and probability space[J]. *Pattern Recognition and Artificial Intelligence*, 2013, 26(6): 552-560.
- [12] Wang S C, Song Z X, Ding H, et al. An improved reinforcement Q-learning method with BP neural networks in robot soccer[C]//4th International Symposium on Computational Intelligence and Design. Piscataway, USA: IEEE, 2011: 177-180.
- [13] Jiang Y L. Task allocation of multiple robotic fishes based on self-organizing map neural network[J]. *Applied Mechanics and Materials*, 2014, 563: 308-311.
- [14] 朱大奇, 曹翔. 多个水下机器人动态任务分配和路径规划的信度自组织算法 [J]. *控制理论与应用*, 2015, 32(6): 762-769.
Zhu D Q, Cao X. An improved self-organizing map method for multiple autonomous underwater vehicle teams in dynamic task assignment and path planning[J]. *Control Theory and Applications*, 2015, 32(6): 762-769.
- [15] Ma Z J, Cui Y Y. Optimal hierarchical allocation in deregulated electricity market under PSP auction mechanism[C]//26th Chinese Control and Decision Conference. Piscataway, USA: IEEE, 2014: 2256-2261.

作者简介:

孙博寒 (1992 -), 男, 硕士生. 研究领域: 多机器人协作, 情感机器人, 机器学习, 数据挖掘.

王浩 (1962 -), 男, 博士, 教授, 博士生导师. 研究领域: 人工智能与机器人, 数据挖掘.

方宝富 (1978 -), 男, 博士, 副教授, 硕士生导师. 研究领域: 多智能体系统, 情感机器人, 机器学习.