



(10) 授权公告号 CN 116452011 B

(45) 授权公告日 2023. 10. 24

(21) 申请号 202310265401.9

G06Q 10/0631 (2023.01)

(22) 申请日 2023.03.14

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 114996856 A, 2022.09.02

申请公布号 CN 116452011 A

CN 115358365 A, 2022.11.18

(43) 申请公布日 2023.07.18

CN 114897267 A, 2022.08.12

(73) 专利权人 中国人民解放军32370部队

CN 109544082 A, 2019.03.29

地址 100042 北京市石景山区石门路258号

CN 112221149 A, 2021.01.15

(72) 发明人 张扬 张俊峰 陈企华 淦家杭

CN 112307622 A, 2021.02.02

周鹏 张煜 张媛媛 王常琳

CN 113128705 A, 2021.07.16

张焕俊 李文哲 高超 田原

CN 115660155 A, 2023.01.31

张晓文

US 2022172264 A1, 2022.06.02

(74) 专利代理机构 北京丰浩知识产权代理事务

钟友武 等. 自主近距离空中敌机的战术动作识别方法. 北京航空航天大学学报. 2007, 第33卷(第09期), 第1056-1059页.

所(普通合伙) 11781

专利代理师 李强

审查员 吴秀萍

(51) Int. Cl.

G06Q 10/0637 (2023.01)

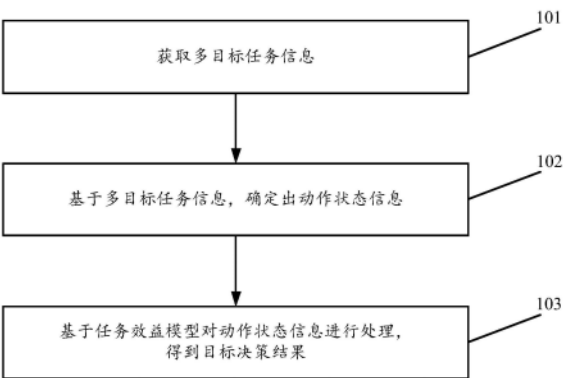
权利要求书3页 说明书11页 附图2页

(54) 发明名称

一种用于智能体决策的数据处理方法及装置

(57) 摘要

本发明公开了一种用于智能体决策的数据处理方法及装置,该方法包括:获取多目标任务信息;多目标任务信息包括第一任务对象信息和第二任务对象信息;第一任务对象信息包括N个第一任务对象;第二任务对象信息包括M个第二任务对象;N不小于M;基于多目标任务信息,确定出动作状态信息;动作状态信息包括M个并列排列的 $N \times N$ 矩阵;基于任务效益模型对动作状态信息进行处理,得到目标决策结果;目标决策结果用于指示第一任务对象对第二任务对象的动作执行。可见,本发明有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。



1. 一种用于智能体决策的数据处理方法,其特征在于,所述方法包括:

获取多目标任务信息;所述多目标任务信息包括第一任务对象信息和第二任务对象信息;所述第一任务对象信息包括N个第一任务对象;所述第二任务对象信息包括M个第二任务对象;所述N不小于所述M;

基于所述多目标任务信息,确定出动作状态信息;所述动作状态信息包括所述M个并列排布的 $N \times N$ 矩阵;

基于任务效益模型对所述动作状态信息进行处理,得到目标决策结果;所述目标决策结果用于指示所述第一任务对象对所述第二任务对象的动作执行;其中,所述基于任务效益模型对所述动作状态信息进行处理,得到目标决策结果,包括:

基于所述动作状态信息,确定出目标动作状态;所述目标动作状态为 $M \times N$ 的矩阵;

获取当前迭代次数;

判断当前迭代次数是否满足迭代阈值,得到迭代判断结果;

当所述迭代判断结果为是时,基于所述目标动作状态和所有历史目标动作状态,确定出目标决策结果;

当所述迭代判断结果为否时,利用任务效益模型对所述目标动作状态进行计算处理,得到动作奖励值;

基于所述目标动作状态和所述动作奖励值,确定出所述目标决策结果;其中,所述利用任务效益模型对所述目标动作状态进行计算处理,得到动作奖励值,包括:

判断所述目标动作状态中的是否存在矩阵元素全部为0的列,得到奖励判断结果;

当所述奖励判断结果为是时,确定动作奖励值为奖励阈值;所述奖励阈值为不大于0的整数;

当所述奖励判断结果为否时,基于任务效益模型对所述目标动作状态进行计算,得到所述动作奖励值;

所述任务效益模型为:

$$R = \max(\sum_{j=1}^N (v_j \sum_{i=1}^M x_{ij})) - \min(\sum_{i=1}^M (c_i \sum_{j=1}^N x_{ij}));$$

其中,R为动作奖励值; v_j 为第j个第二任务对象的综合价值; c_i 为第i个第一任务对象的成本; x_{ij} 为目标动作状态中的状态动作;其中,所述基于所述目标动作状态和所述动作奖励值,确定出所述目标决策结果,包括:

判断所述动作奖励值是否满足最优性约束,得到最优判断结果;所述最优性约束表征动作奖励值优于预设的奖励阈值;

当所述最优判断结果为否时,对所述目标动作状态进行更新,并出发执行所述获取当前迭代次数;

当所述最优判断结果为是时,确定所述目标动作状态为所述目标决策结果。

2. 根据权利要求1所述的用于智能体决策的数据处理方法,其特征在于,所述对所述目标动作状态进行更新,包括:

响应于所述目标动作状态的仿真动作执行,得到当前任务状态;

利用所述当前任务状态与所述动作状态信息之间的状态匹配关系,确定出待用动作状态;

利用优解模型对所述待用动作状态和所有所述历史目标动作状态进行计算,得到当前动作状态;

所述优解模型为:

$$A = \begin{cases} \operatorname{argmax}(Q) & \operatorname{random}() > \varepsilon \\ Q_x & \operatorname{random}() \leq \varepsilon \end{cases};$$

其中,A为当前动作状态;Q为所有所述历史目标动作状态; Q_x 为所述待用动作状态; ε 为优解系数; $\operatorname{random}()$ 为所述待用动作状态对应的状态概率值;

利用更新模型将所述当前动作状态替换为新的目标动作状态;

所述更新模型为:

$$Q(s1, a1) \leftarrow Q(s0, a0) + \alpha [R + \gamma \max_a Q(s', a) - Q(s0, a0)];$$

其中, $Q(s1, a1)$ 为新的目标动作状态; $Q(s0, a0)$ 为更新前的目标动作状态; α 为校正系数; γ 为折扣系数; $Q(s', a)$ 为所述当前动作状态。

3. 根据权利要求2所述的用于智能体决策的数据处理方法,其特征在于,所述当前动作状态为 $M \times N$ 的矩阵;所述当前动作状态的矩阵元素为0,和/或,1;

在所述利用优解模型对所述待用动作状态和所有所述历史目标动作状态进行计算,得到当前动作状态之后,所述方法还包括:

计算所述当前动作状态中所有所述矩阵元素的和,得到元素和值;

判断所述元素和值是否等于所述M,得到元素判断结果;

当所述元素判断结果为否时,触发执行所述利用所述当前任务状态与所述动作状态信息之间的状态匹配关系,确定出待用动作状态;

当所述元素判断结果为是时,触发执行所述利用优解模型对所述待用动作状态和所有所述历史目标动作状态进行计算,得到当前动作状态。

4. 根据权利要求1所述的用于智能体决策的数据处理方法,其特征在于,所述基于所述多目标任务信息,确定出动作状态信息,包括:

对所述第二任务对象信息中的第二任务对象按序进行编排,得到动作对象序列;

对于所述第一任务对象信息中的任一所述第一任务对象,依据所述动作对象序列,生成M个状态动作向量;任意2个所述状态动作向量是不相一致的;所述状态动作向量中所有向量元素之和不大于所述N;所述状态动作向量的向量元素为0,和/或,1;

按所述状态动作向量的生成顺序,依序排列所述M个所述状态动作向量,得到状态动作矩阵;

构建所有所述状态动作矩阵进行并列对应关系,得到动作状态信息;所述并列对应关系表征任意2个所述状态动作矩阵中同一位置的矩阵元素不同时为1。

5. 一种用于智能体决策的数据处理装置,其特征在于,所述装置包括:

获取模块,用于获取多目标任务信息;所述多目标任务信息包括第一任务对象信息和第二任务对象信息;所述第一任务对象信息包括N个第一任务对象;所述第二任务对象信息包括M个第二任务对象;所述N不小于所述M;

确定模块,用于基于所述多目标任务信息,确定出动作状态信息;所述动作状态信息包括所述M个并列排布的 $N \times N$ 矩阵;

得到模块,用于基于任务效益模型对所述动作状态信息进行处理,得到目标决策结果;所述目标决策结果用于指示所述第一任务对象对所述第二任务对象的动作执行;

其中,所述基于任务效益模型对所述动作状态信息进行处理,得到目标决策结果,包括:

基于所述动作状态信息,确定出目标动作状态;所述目标动作状态为 $M \times N$ 的矩阵;

获取当前迭代次数;

判断当前迭代次数是否满足迭代阈值,得到迭代判断结果;

当所述迭代判断结果为是时,基于所述目标动作状态和所有历史目标动作状态,确定出目标决策结果;

当所述迭代判断结果为否时,利用任务效益模型对所述目标动作状态进行计算处理,得到动作奖励值;

基于所述目标动作状态和所述动作奖励值,确定出所述目标决策结果;其中,所述利用任务效益模型对所述目标动作状态进行计算处理,得到动作奖励值,包括:

判断所述目标动作状态中的是否存在矩阵元素全部为0的列,得到奖励判断结果;

当所述奖励判断结果为是时,确定动作奖励值为奖励阈值;所述奖励阈值为不大于0的整数;

当所述奖励判断结果为否时,基于任务效益模型对所述目标动作状态进行计算,得到所述动作奖励值;

所述任务效益模型为:

$$R = \max(\sum_{j=1}^N (v_j \sum_{i=1}^M x_{ij})) - \min(\sum_{i=1}^M (c_i \sum_{j=1}^N x_{ij}));$$

其中, R 为动作奖励值; v_j 为第 j 个第二任务对象的综合价值; c_i 为第 i 个第一任务对象的成本; x_{ij} 为目标动作状态中的状态动作;其中,所述基于所述目标动作状态和所述动作奖励值,确定出所述目标决策结果,包括:

判断所述动作奖励值是否满足最优性约束,得到最优判断结果;所述最优性约束表征动作奖励值优于预设的奖励阈值;

当所述最优判断结果为否时,对所述目标动作状态进行更新,并出发执行所述获取当前迭代次数;

当所述最优判断结果为是时,确定所述目标动作状态为所述目标决策结果。

6.一种用于智能体决策的数据处理装置,其特征在于,所述装置包括:

存储有可执行程序代码的存储器;

与所述存储器耦合的处理器;

所述处理器调用所述存储器中存储的所述可执行程序代码,执行如权利要求1-4任一项所述的用于智能体决策的数据处理方法。

7.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质存储有计算机指令,所述计算机指令被调用时,用于执行如权利要求1-4任一项所述的用于智能体决策的数据处理方法。

一种用于智能体决策的数据处理方法及装置

技术领域

[0001] 本发明涉及仿真技术领域,尤其涉及一种用于智能体决策的数据处理方法及装置。

背景技术

[0002] 随着武器技术的发展,传统的攻防对抗场景呈现不断复杂化的趋势。传统一对一的攻防对抗问题也难以满足复杂战场智能化的发展趋势。在各种新的研究中,对抗双方不断采用新的博弈策略从而取得战场优势。其中由目标主动防御引发的三体对抗问题成为研究的热点。一个多对多的博弈问题不能简单地分解为多个一对一单智能体博弈问题。因此,提供一种用于智能体决策的数据处理方法及装置,以实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

发明内容

[0003] 本发明所要解决的技术问题在于,提供一种用于智能体决策的数据处理方法及装置有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0004] 为了解决上述技术问题,本发明实施例第一方面公开了一种用于智能体决策的数据处理方法,所述方法包括:

[0005] 获取多目标任务信息;所述多目标任务信息包括第一任务对象信息和第二任务对象信息;所述第一任务对象信息包括N个第一任务对象;所述第二任务对象信息包括M个第二任务对象;所述N不小于所述M;

[0006] 基于所述多目标任务信息,确定出动作状态信息;所述动作状态信息包括所述M个并列排布的 $N \times N$ 矩阵;

[0007] 基于任务效益模型对所述动作状态信息进行处理,得到目标决策结果;所述目标决策结果用于指示所述第一任务对象对所述第二任务对象的动作执行。

[0008] 作为一种可选的实施方式,在本发明实施例第一方面中,所述基于任务效益模型对所述动作状态信息进行处理,得到目标决策结果,包括:

[0009] 基于所述动作状态信息,确定出目标动作状态;所述目标动作状态为 $M \times N$ 的矩阵;

[0010] 获取当前迭代次数;

[0011] 判断当前迭代次数是否满足迭代阈值,得到迭代判断结果;

[0012] 当所述迭代判断结果为是时,基于所述目标动作状态和所有历史目标动作状态,确定出目标决策结果;

[0013] 当所述迭代判断结果为否时,利用任务效益模型对所述目标动作状态进行计算处理,得到动作奖励值;

[0014] 基于所述目标动作状态和所述动作奖励值,确定出所述目标决策结果。

[0015] 作为一种可选的实施方式,在本发明实施例第一方面中,所述利用任务效益模型

对所述目标动作状态进行计算处理,得到动作奖励值,包括:

[0016] 判断所述目标动作状态中的是否存在矩阵元素全部为0的列,得到奖励判断结果;

[0017] 当所述奖励判断结果为是时,确定动作奖励值为奖励阈值;所述奖励阈值为不大于0的整数;

[0018] 当所述奖励判断结果为否时,基于任务效益模型对所述目标动作状态进行计算,得到所述动作奖励值;

[0019] 所述任务效益模型为:

$$[0020] \quad R = \max(\sum_{j=1}^N (v_j \sum_{i=1}^M x_{ij})) - \min(\sum_{i=1}^M (c_i \sum_{j=1}^N x_{ij}));$$

[0021] 其中,R为动作奖励值; v_j 为第j个第二任务对象的综合价值; c_i 为第i个第一任务对象的成本; x_{ij} 为目标动作状态中的状态动作。

[0022] 作为一种可选的实施方式,在本发明实施例第一方面中,所述基于所述目标动作状态和所述动作奖励值,确定出所述目标决策结果,包括:

[0023] 判断所述动作奖励值是否满足最优性约束,得到最优判断结果;所述最优性约束表征动作奖励值优于预设的奖励阈值;

[0024] 当所述最优判断结果为否时,对所述目标动作状态进行更新,并出发执行所述获取当前迭代次数;

[0025] 当所述最优判断结果为是时,确定所述目标动作状态为所述目标决策结果。

[0026] 作为一种可选的实施方式,在本发明实施例第一方面中,所述对所述目标动作状态进行更新,包括:

[0027] 响应于所述目标动作状态的仿真动作执行,得到当前任务状态;

[0028] 利用所述当前任务状态与所述动作状态信息之间的状态匹配关系,确定出待用动作状态;

[0029] 利用优解模型对所述待用动作状态和所有所述历史目标动作状态进行计算,得到当前动作状态;

[0030] 所述优解模型为:

$$[0031] \quad A = \begin{cases} \operatorname{argmax}(Q) & \operatorname{random}() > \varepsilon \\ Q_x & \operatorname{random}() \leq \varepsilon \end{cases};$$

[0032] 其中,A为当前动作状态;Q为所有所述历史目标动作状态; Q_x 为所述待用动作状态; ε 为优解系数; $\operatorname{random}()$ 为所述待用动作状态对应的状态概率值;

[0033] 利用更新模型将所述当前动作状态替换为新的目标动作状态;

[0034] 所述更新模型为:

$$[0035] \quad Q(s1, a1) \leftarrow Q(s0, a0) + \alpha [R + \gamma \max_a Q(s', a) - Q(s0, a0)];$$

[0036] 其中, $Q(s1, a1)$ 为新的目标动作状态; $Q(s0, a0)$ 为更新前的目标动作状态; α 为校正系数; γ 为折扣系数; $Q(s', a)$ 为当前动作状态。

[0037] 作为一种可选的实施方式,在本发明实施例第一方面中,所述当前动作状态为 $M \times N$ 的矩阵;所述当前动作状态的矩阵元素为0,和/或,1;

[0038] 在所述利用优解模型对所述待用动作状态和所有所述历史目标动作状态进行计算,得到当前动作状态之后,所述方法还包括:

- [0039] 计算所述当前动作状态中所有所述矩阵元素的和,得到元素和值;
- [0040] 判断所述元素和值是否等于所述M,得到元素判断结果;
- [0041] 当所述元素判断结果为否时,触发执行所述利用所述当前任务状态与所述动作状态信息之间的状态匹配关系,确定出待用动作状态;
- [0042] 当所述元素判断结果为是时,触发执行所述利用优解模型对所述待用动作状态和所有所述历史目标动作状态进行计算,得到当前动作状态。
- [0043] 作为一种可选的实施方式,在本发明实施例第一方面中,所述基于所述多目标任务信息,确定出动作状态信息,包括:
- [0044] 对所述第二任务对象信息中的第二任务对象按序进行编排,得到动作对象序列;
- [0045] 对于所述第一任务对象信息中的任一所述第一任务对象,依据所述动作对象序列,生成M个状态动作向量;任意2个所述状态动作向量是不相一致的;所述状态动作向量中所有向量元素之和不大于所述N;所述状态动作向量的向量元素为0,和/或,1;
- [0046] 按所述状态动作向量的生成顺序,依序排列所述M个所述状态动作向量,得到状态动作矩阵;
- [0047] 构建所有所述状态动作矩阵进行并列对应关系,得到动作状态信息;所述并列对应关系表征任意2个所述状态动作矩阵中同一位置的矩阵元素不同时为1。
- [0048] 本发明实施例第二方面公开了一种用于智能体决策的数据处理装置,装置包括:
- [0049] 获取模块,用于获取多目标任务信息;所述多目标任务信息包括第一任务对象信息和第二任务对象信息;所述第一任务对象信息包括N个第一任务对象;所述第二任务对象信息包括M个第二任务对象;所述N不小于所述M;
- [0050] 确定模块,用于基于所述多目标任务信息,确定出动作状态信息;所述动作状态信息包括所述M个并列排布的 $N \times N$ 矩阵;
- [0051] 得到模块,用于基于任务效益模型对所述动作状态信息进行处理,得到目标决策结果;所述目标决策结果用于指示所述第一任务对象对所述第二任务对象的动作执行。
- [0052] 本发明第三方面公开了另一种用于智能体决策的数据处理装置,所述装置包括:
- [0053] 存储有可执行程序代码的存储器;
- [0054] 与所述存储器耦合的处理器;
- [0055] 所述处理器调用所述存储器中存储的所述可执行程序代码,执行本发明实施例第一方面公开的用于智能体决策的数据处理方法中的部分或全部步骤。
- [0056] 本发明第四方面公开了一种计算机可读存储介质,所述计算机可读存储介质存储有计算机指令,所述计算机指令被调用时,用于执行本发明实施例第一方面公开的用于智能体决策的数据处理方法中的部分或全部步骤。
- [0057] 与现有技术相比,本发明实施例具有以下有益效果:
- [0058] 本发明实施例中,获取多目标任务信息;多目标任务信息包括第一任务对象信息和第二任务对象信息;第一任务对象信息包括N个第一任务对象;第二任务对象信息包括M个第二任务对象;N不小于M;基于多目标任务信息,确定出动作状态信息;动作状态信息包括M个并列排布的 $N \times N$ 矩阵;基于任务效益模型对动作状态信息进行处理,得到目标决策结果;目标决策结果用于指示第一任务对象对第二任务对象的动作执行。可见,本发明有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

附图说明

[0059] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0060] 图1是本发明实施例公开的一种用于智能体决策的数据处理方法的流程示意图;

[0061] 图2是本发明实施例公开的一种用于智能体决策的数据处理装置的结构示意图;

[0062] 图3是本发明实施例公开的另一种用于智能体决策的数据处理装置的结构示意图。

具体实施方式

[0063] 为了使本技术领域的人员更好地理解本发明方案,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0064] 本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别不同对象,而不是用于描述特定顺序。此外,术语“包括”和“具有”以及它们任何变形,意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、装置、产品或设备没有限定于已列出的步骤或单元,而是可选地还包括没有列出的步骤或单元,或可选地还包括对于这些过程、方法、产品或设备固有的其他步骤或单元。

[0065] 在本文中提及“实施例”意味着,结合实施例描述的特定特征、结构或特性可以包含在本发明的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例,也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是,本文所描述的实施例可以与其它实施例相结合。

[0066] 本发明公开了一种用于智能体决策的数据处理方法及装置有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。以下分别进行详细说明。

[0067] 实施例一

[0068] 请参阅图1,图1是本发明实施例公开的一种用于智能体决策的数据处理方法的流程示意图。其中,图1所描述的用于智能体决策的数据处理方法应用于仿真决策系统中,如用于智能体决策的数据处理管理的本地服务器或云端服务器等,本发明实施例不做限定。如图1所示,该用于智能体决策的数据处理方法可以包括以下操作:

[0069] 101、获取多目标任务信息。

[0070] 本发明实施例中,多目标任务信息包括第一任务对象信息和第二任务对象信息。

[0071] 本发明实施例中,第一任务对象信息包括N个第一任务对象。

[0072] 本发明实施例中,第二任务对象信息包括M个第二任务对象。

[0073] 本发明实施例中,N不小于M。

[0074] 102、基于多目标任务信息,确定出动作状态信息。

[0075] 本发明实施例中,动作状态信息包括M个并列排布的 $N \times N$ 矩阵。

- [0076] 103、基于任务效益模型对动作状态信息进行处理,得到目标决策结果。
- [0077] 本发明实施例中,目标决策结果用于指示第一任务对象对第二任务对象的动作执行。
- [0078] 需要说明的是,第一任务对象可以为作战仿真中的主动动作的仿真主体。具体的,其可以为战斗机,和/或,防御导弹。
- [0079] 需要说明的是,第二任务对象可以为作战仿真中的来袭动作的仿真主体。具体的,其可以为来袭导弹。
- [0080] 需要说明的是,上述并列排布的 $N \times N$ 矩阵表征第一任务对象对第二任务对象的可能并行动作。
- [0081] 可见,实施本发明实施例所描述的用于智能体决策的数据处理方法有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。
- [0082] 在一个可选的实施例中,上述基于任务效益模型对动作状态信息进行处理,得到目标决策结果,包括:
- [0083] 基于动作状态信息,确定出目标动作状态;目标动作状态为 $M \times N$ 的矩阵;
- [0084] 获取当前迭代次数;
- [0085] 判断当前迭代次数是否满足迭代阈值,得到迭代判断结果;
- [0086] 当迭代判断结果为是时,基于目标动作状态和所有历史目标动作状态,确定出目标决策结果;
- [0087] 当迭代判断结果为否时,利用任务效益模型对目标动作状态进行计算处理,得到动作奖励值;
- [0088] 基于目标动作状态和动作奖励值,确定出目标决策结果。
- [0089] 需要说明的是,上述当前迭代次数表征多目标博弈决策的循环次数,每次迭代之后次数增加一次。
- [0090] 在该可选的实施例中,作为一种可选的实施方式,上述基于动作状态信息,确定出目标动作状态,包括:
- [0091] 获取当前任务状态;
- [0092] 基于当前任务状态与动作状态信息中状态动作矩阵的状态匹配关系,确定出 M 个目标状态动作向量;
- [0093] 将 M 个目标状态动作向量依据动作对象序列进行编码,得到目标动作状态。
- [0094] 需要说明的是,上述状态匹配关系可以是预先设定的,也可以是根据历史数据训练得到的,本发明实施例不做限定。
- [0095] 需要说明的是,上述当前任务状态表征第一任务对象与第二任务对象的动作匹配情况。
- [0096] 在该可选的实施例中,作为一种可选的实施方式,基于目标动作状态和所有历史目标动作状态,确定出目标决策结果,包括:
- [0097] 对目标动作状态和所有历史目标动作状态按奖励函数值从大到小进行排序,得到动作状态序列;
- [0098] 确定动作状态序列中排序第一的元素为目标决策结果。
- [0099] 可见,实施本发明实施例所描述的用于智能体决策的数据处理方法有利于实现每

个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0100] 在另一个可选的实施例中,利用任务效益模型对目标动作状态进行计算处理,得到动作奖励值,包括:

[0101] 判断目标动作状态中的是否存在矩阵元素全部为0的列,得到奖励判断结果;

[0102] 当奖励判断结果为是时,确定动作奖励值为奖励阈值;奖励阈值为不大于0的整数;

[0103] 当奖励判断结果为否时,基于任务效益模型对目标动作状态进行计算,得到动作奖励值;

[0104] 任务效益模型为:

$$[0105] \quad R = \max(\sum_{j=1}^N (v_j \sum_{i=1}^M x_{ij})) - \min(\sum_{i=1}^M (c_i \sum_{j=1}^N x_{ij}));$$

[0106] 其中,R为动作奖励值; v_j 为第j个第二任务对象的综合价值; c_i 为第i个第一任务对象的成本; x_{ij} 为目标动作状态中的状态动作。

[0107] 需要说明的是,上述目标动作状态中存在矩阵元素全部为0的列表征第一任务对象没有全部被分配动作作用于第二任务对象。

[0108] 需要说明的是,通过计算动作奖励值来判断当前的目标动作状态是否满足最优性约束,来指导对目标动作状态的优化选择,从而避免在多目标决策过程中生成不符合约束条件的情况出现,以提高数据处理的效率和精度。

[0109] 可见,实施本发明实施例所描述的用于智能体决策的数据处理方法有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0110] 在又一个可选的实施例中,基于目标动作状态和动作奖励值,确定出目标决策结果,包括:

[0111] 判断动作奖励值是否满足最优性约束,得到最优判断结果;最优性约束表征动作奖励值优于预设的奖励阈值;

[0112] 当最优判断结果为否时,对目标动作状态进行更新,并出发执行获取当前迭代次数;

[0113] 当最优判断结果为是时,确定目标动作状态为目标决策结果。

[0114] 需要说明的是,上述奖励阈值可以是预先设定的,也可以是动态变化的,本发明实施例不做限定。

[0115] 可见,实施本发明实施例所描述的用于智能体决策的数据处理方法有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0116] 在又一个可选的实施例中,对目标动作状态进行更新,包括:

[0117] 响应于目标动作状态的仿真动作执行,得到当前任务状态;

[0118] 利用当前任务状态与动作状态信息之间的状态匹配关系,确定出待用动作状态;

[0119] 利用优解模型对待用动作状态和所有历史目标动作状态进行计算,得到当前动作状态;

[0120] 优解模型为:

$$[0121] \quad A = \begin{cases} \operatorname{argmax}(Q) & \operatorname{random}() > \varepsilon \\ Q_x & \operatorname{random}() \leq \varepsilon \end{cases};$$

[0122] 其中,A为当前动作状态;Q为所有历史目标动作状态; Q_x 为待用动作状态; ϵ 为优解系数;random()为待用动作状态对应的状态概率值;

[0123] 利用更新模型将当前动作状态替换为新的目标动作状态;

[0124] 更新模型为:

$$[0125] \quad Q(s1, a1) \leftarrow Q(s0, a0) + \alpha[R + \gamma \max_a Q(s', a) - Q(s0, a0)];$$

[0126] 其中,Q(s1,a1)为新的目标动作状态;Q(s0,a0)为更新前的目标动作状态; α 为校正系数; γ 为折扣系数;Q(s',a)为当前动作状态。

[0127] 需要说明的是,上述响应于目标动作状态的仿真动作执行是基于目标动作状态执行第一任务对象对第二任务对象的动作,以得到第一任务对象和第二任务对象对应的动作状态情况,即当前任务状态。

[0128] 可见,实施本发明实施例所描述的用于智能体决策的数据处理方法有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0129] 在一个可选的实施例中,上述当前动作状态为M×N的矩阵;当前动作状态的矩阵元素为0,和/或,1;

[0130] 在利用优解模型对待用动作状态和所有历史目标动作状态进行计算,得到当前动作状态之后,方法还包括:

[0131] 计算当前动作状态中所有矩阵元素的和,得到元素和值;

[0132] 判断元素和值是否等于M,得到元素判断结果;

[0133] 当元素判断结果为否时,触发执行利用当前任务状态与动作状态信息之间的状态匹配关系,确定出待用动作状态;

[0134] 当元素判断结果为是时,触发执行利用优解模型对待用动作状态和所有历史目标动作状态进行计算,得到当前动作状态。

[0135] 可见,实施本发明实施例所描述的用于智能体决策的数据处理方法有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0136] 在另一个可选的实施例中,基于多目标任务信息,确定出动作状态信息,包括:

[0137] 对第二任务对象信息中的第二任务对象按序进行编排,得到动作对象序列;

[0138] 对于第一任务对象信息中的任一第一任务对象,依据动作对象序列,生成M个状态动作向量;任意2个状态动作向量是不相一致的;状态动作向量中所有向量元素之和不大于N;状态动作向量的向量元素为0,和/或,1;

[0139] 按状态动作向量的生成顺序,依序排列M个状态动作向量,得到状态动作矩阵;

[0140] 构建所有状态动作矩阵进行并列对应关系,得到动作状态信息;并列对应关系表征任意2个状态动作矩阵中同一位置的矩阵元素不同时为1。

[0141] 需要说明的,上述动作对象序列是将第二任务对象进行标号,以便于后续对第二任务对象的有序处理。

[0142] 需要说明的,上述通过对于第一任务对象信息中的任一第一任务对象,依据动作对象序列,生成M个状态动作向量可将N个第一任务对象对M个第二任务对象的所有可能动作。举例来说,当第一任务对象为4枚拦截弹,第二任务对象为2枚来袭导弹,则M个状态动作向量可如下表所示:

[0143]	可能性	状态动作向量	可能性	状态动作向量
	1	[0 0 0 0]	9	[1 0 0 0]
	2	[0 0 0 1]	10	[1 0 0 1]
	3	[0 0 1 0]	11	[1 0 1 0]
	4	[0 0 1 1]	12	[1 0 1 1]
	5	[0 1 0 0]	13	[1 1 0 0]
	6	[0 1 0 1]	14	[1 1 0 1]
	7	[0 1 1 0]	15	[1 1 1 0]
	8	[0 1 1 1]	16	[1 1 1 1]

[0144] 可见,实施本发明实施例所描述的用于智能体决策的数据处理方法有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0145] 实施例二

[0146] 请参阅图2,图2是本发明实施例公开的一种用于智能体决策的数据处理装置的结构示意图。其中,图2所描述的装置能够应用于仿真决策系统中,如用于智能体决策的数据处理管理的本地服务器或云端服务器等,本发明实施例不做限定。如图2所示,该装置可以包括:

[0147] 获取模块,用于获取多目标任务信息;多目标任务信息包括第一任务对象信息和第二任务对象信息;第一任务对象信息包括N个第一任务对象;第二任务对象信息包括M个第二任务对象;N不小于M;

[0148] 确定模块,用于基于多目标任务信息,确定出动作状态信息;动作状态信息包括M个并列排布的 $N \times N$ 矩阵;

[0149] 得到模块,用于基于任务效益模型对动作状态信息进行处理,得到目标决策结果;目标决策结果用于指示第一任务对象对第二任务对象的动作执行。

[0150] 可见,实施图2所描述的用于智能体决策的数据处理装置有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0151] 在另一个可选的实施例中,如图2所示,得到模块基于任务效益模型对动作状态信息进行处理,得到目标决策结果,包括:

[0152] 基于动作状态信息,确定出目标动作状态;目标动作状态为 $M \times N$ 的矩阵;

[0153] 获取当前迭代次数;

[0154] 判断当前迭代次数是否满足迭代阈值,得到迭代判断结果;

[0155] 当迭代判断结果为是时,基于目标动作状态和所有历史目标动作状态,确定出目标决策结果;

[0156] 当迭代判断结果为否时,利用任务效益模型对目标动作状态进行计算处理,得到动作奖励值;

[0157] 基于目标动作状态和动作奖励值,确定出目标决策结果。

[0158] 可见,实施图2所描述的用于智能体决策的数据处理装置有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0159] 在又一个可选的实施例中,如图2所示,得到模块利用任务效益模型对目标动作状态进行计算处理,得到动作奖励值,包括:

[0160] 判断目标动作状态中的是否存在矩阵元素全部为0的列,得到奖励判断结果;

[0161] 当奖励判断结果为是时,确定动作奖励值为奖励阈值;奖励阈值为不大于0的整数;

[0162] 当奖励判断结果为否时,基于任务效益模型对目标动作状态进行计算,得到动作奖励值;

[0163] 任务效益模型为:

$$[0164] \quad R = \max(\sum_{j=1}^N (v_j \sum_{i=1}^M x_{ij})) - \min(\sum_{i=1}^M (c_i \sum_{j=1}^N x_{ij}));$$

[0165] 其中,R为动作奖励值; v_j 为第j个第二任务对象的综合价值; c_i 为第i个第一任务对象的成本; x_{ij} 为目标动作状态中的状态动作。

[0166] 可见,实施图2所描述的用于智能体决策的数据处理装置有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0167] 在又一个可选的实施例中,如图2所示,得到模块基于目标动作状态和动作奖励值,确定出目标决策结果,包括:

[0168] 判断动作奖励值是否满足最优性约束,得到最优判断结果;最优性约束表征动作奖励值优于预设的奖励阈值;

[0169] 当最优判断结果为否时,对目标动作状态进行更新,并出发执行获取当前迭代次数;

[0170] 当最优判断结果为是时,确定目标动作状态为目标决策结果。

[0171] 可见,实施图2所描述的用于智能体决策的数据处理装置有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0172] 在又一个可选的实施例中,如图2所示,得到模块对目标动作状态进行更新,包括:

[0173] 响应于目标动作状态的仿真动作执行,得到当前任务状态;

[0174] 利用当前任务状态与动作状态信息之间的状态匹配关系,确定出待用动作状态;

[0175] 利用优解模型对待用动作状态和所有历史目标动作状态进行计算,得到当前动作状态;

[0176] 优解模型为:

$$[0177] \quad A = \begin{cases} \operatorname{argmax}(Q) & \operatorname{random}() > \varepsilon \\ Q_x & \operatorname{random}() \leq \varepsilon \end{cases}$$

[0178] 其中,A为当前动作状态;Q为所有历史目标动作状态; Q_x 为待用动作状态; ε 为优解系数; $\operatorname{random}()$ 为待用动作状态对应的状态概率值;

[0179] 利用更新模型将当前动作状态替换为新的目标动作状态;

[0180] 更新模型为:

$$[0181] \quad Q(s1, a1) \leftarrow Q(s0, a0) + \alpha [R + \gamma \max_a Q(s', a) - Q(s0, a0)];$$

[0182] 其中, $Q(s1, a1)$ 为新的目标动作状态; $Q(s0, a0)$ 为更新前的目标动作状态; α 为校正系数; γ 为折扣系数; $Q(s', a)$ 为当前动作状态。

[0183] 可见,实施图2所描述的用于智能体决策的数据处理装置有利于实现每个子单位

的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0184] 在又一个可选的实施例中,如图2所示,当前动作状态为 $M \times N$ 的矩阵;当前动作状态的矩阵元素为0,和/或,1;

[0185] 在得到模块利用优解模型对待用动作状态和所有历史目标动作状态进行计算,得到当前动作状态之后,得到模块还用于:

[0186] 计算当前动作状态中所有矩阵元素的和,得到元素和值;

[0187] 判断元素和值是否等于 M ,得到元素判断结果;

[0188] 当元素判断结果是否为否时,触发执行利用当前任务状态与动作状态信息之间的状态匹配关系,确定出待用动作状态;

[0189] 当元素判断结果为是时,触发执行利用优解模型对待用动作状态和所有历史目标动作状态进行计算,得到当前动作状态。

[0190] 可见,实施图2所描述的用于智能体决策的数据处理装置有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0191] 在又一个可选的实施例中,如图2所示,确定模块基于多目标任务信息,确定出动作状态信息,包括:

[0192] 对第二任务对象信息中的第二任务对象按序进行编排,得到动作对象序列;

[0193] 对于第一任务对象信息中的任一第一任务对象,依据动作对象序列,生成 M 个状态动作向量;任意2个状态动作向量是不相一致的;状态动作向量中所有向量元素之和不大于 N ;状态动作向量的向量元素为0,和/或,1;

[0194] 按状态动作向量的生成顺序,依序排列 M 个状态动作向量,得到状态动作矩阵;

[0195] 构建所有状态动作矩阵进行并列对应关系,得到动作状态信息;并列对应关系表征任意2个状态动作矩阵中同一位置的矩阵元素不同时为1。

[0196] 可见,实施图2所描述的用于智能体决策的数据处理装置有利于实现每个子单位的最优目标决策,提高用于智能体决策的数据处理的效率和精准度。

[0197] 实施例三

[0198] 请参阅图3,图3是本发明实施例公开的又一种用于智能体决策的数据处理装置的结构示意图。其中,图3所描述的装置能够应用于仿真决策系统中,如用于智能体决策的数据处理管理的本地服务器或云端服务器等,本发明实施例不做限定。如图3所示,该装置可以包括:

[0199] 存储有可执行程序代码的存储器301;

[0200] 与存储器301耦合的处理器302;

[0201] 处理器302调用存储器301中存储的可执行程序代码,用于执行实施例一所描述的用于智能体决策的数据处理方法中的步骤。

[0202] 实施例四

[0203] 本发明实施例公开了一种计算机可读存储介质,其存储用于电子数据交换的计算机程序,其中,该计算机程序使得计算机执行实施例一所描述的用于智能体决策的数据处理方法中的步骤。

[0204] 实施例五

[0205] 本发明实施例公开了一种计算机程序产品,该计算机程序产品包括存储了计算机

程序的非瞬时性计算机可读存储介质,且该计算机程序可操作来使计算机执行实施例一所描述的用于智能体决策的数据处理方法中的步骤。

[0206] 以上所描述的装置实施例仅是示意性的,其中作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显示的部件可以是或者也可以不是物理模块,即可以位于一个地方,或者也可以分布到多个网络模块上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下,即可以理解并实施。

[0207] 通过以上的实施例的具体描述,本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件。基于这样的理解,上述技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品可以存储在计算机可读存储介质中,存储介质包括只读存储器(Read-Only Memory,ROM)、随机存储器(Random Access Memory,RAM)、可编程只读存储器(Programmable Read-only Memory,PROM)、可擦除可编程只读存储器(Erasable Programmable Read Only Memory,EPR0M)、一次可编程只读存储器(One-time Programmable Read-Only Memory,OTPR0M)、电子抹除式可复写只读存储器(Electrically-Erasable Programmable Read-Only Memory,EEPROM)、只读光盘(Compact Disc Read-Only Memory,CD-ROM)或其他光盘存储器、磁盘存储器、磁带存储器、或者能够用于携带或存储数据的计算机可读的任何其他介质。

[0208] 最后应说明的是:本发明实施例公开的一种用于智能体决策的数据处理方法及装置所揭露的仅为本发明较佳实施例而已,仅用于说明本发明的技术方案,而非对其限制;尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解;其依然可以对前述各项实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或替换,并不使相应的技术方案的本质脱离本发明各项实施例技术方案的精神和范围。

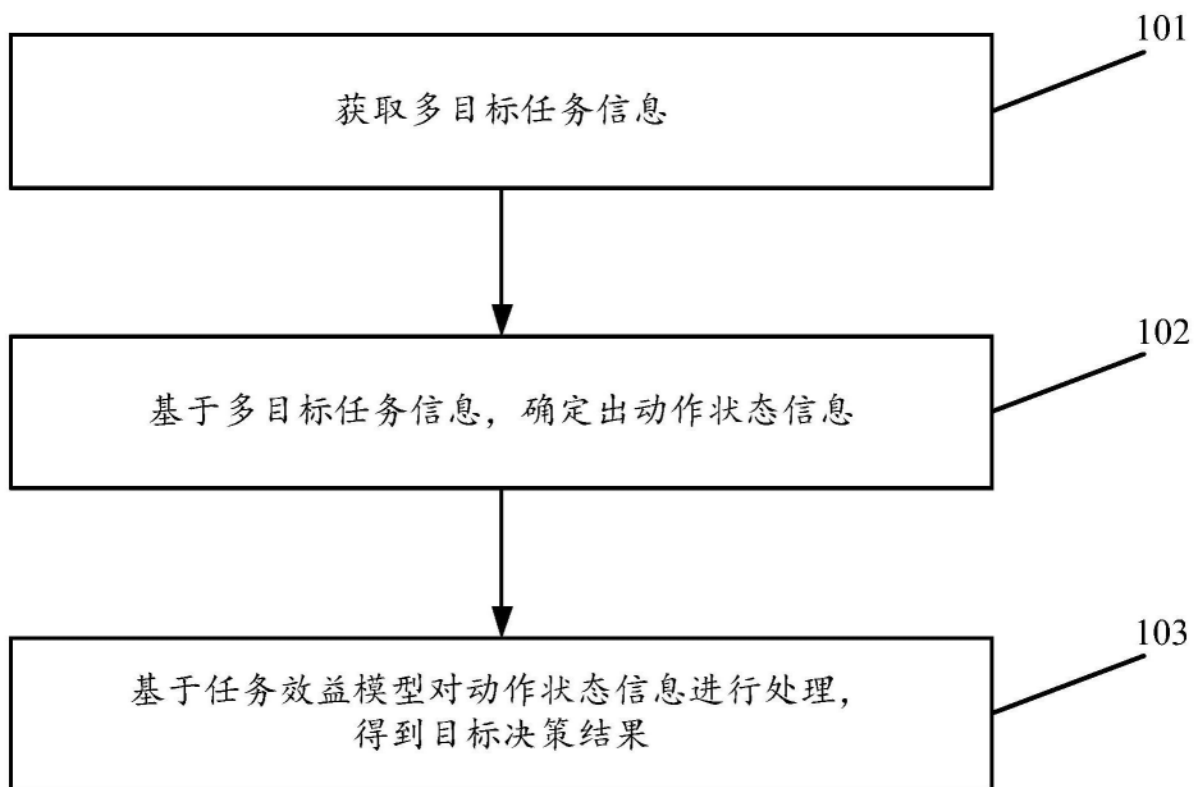


图1

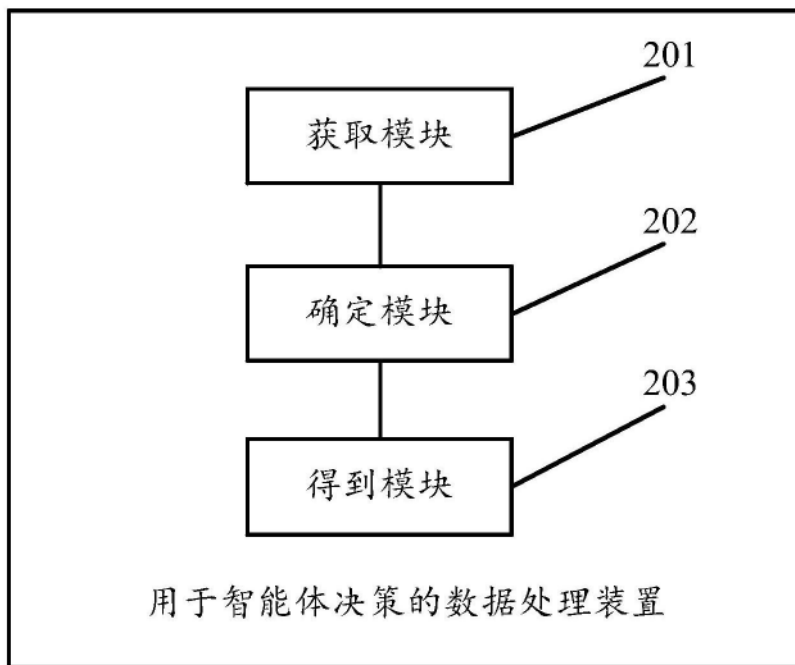


图2

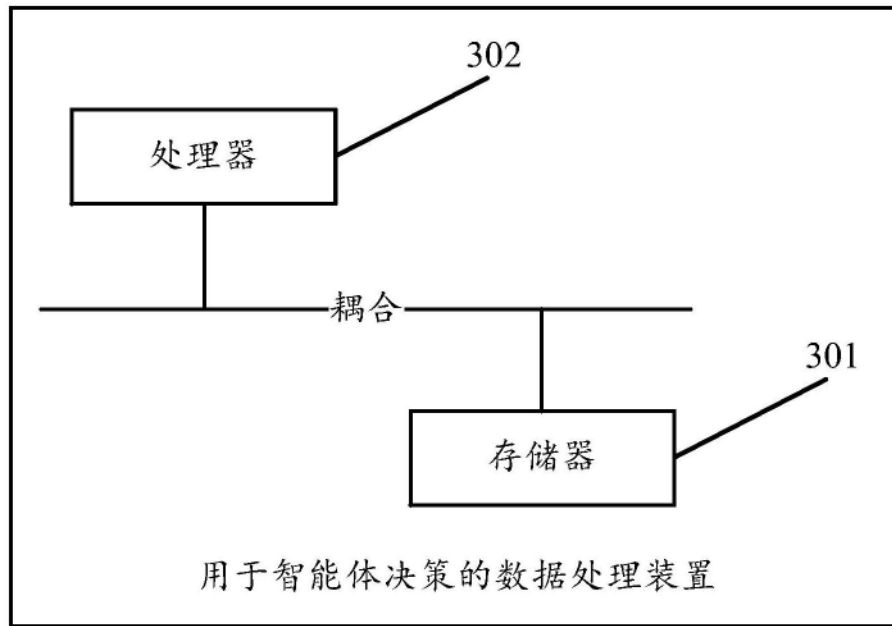


图3