



(12) 发明专利申请

(10) 申请公布号 CN 117314582 A

(43) 申请公布日 2023. 12. 29

(21) 申请号 202311399922.X

(22) 申请日 2023.10.26

(71) 申请人 江西理工大学

地址 341000 江西省赣州市客家大道1958号

(72) 发明人 刘正超 潘春荣 罗鑫 张震 许流洋

(51) Int. Cl.

G06Q 30/0601 (2023.01)

G06N 3/084 (2023.01)

G06N 3/092 (2023.01)

G06Q 50/04 (2012.01)

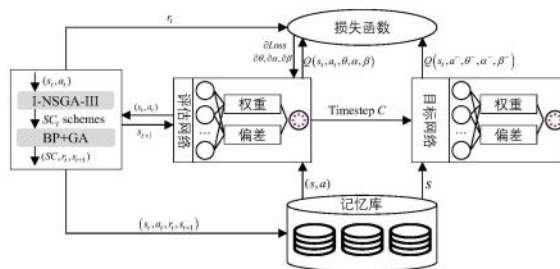
权利要求书2页 说明书5页 附图2页

(54) 发明名称

基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法

(57) 摘要

本发明公开了一种基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法,包括以下步骤:将复杂制造任务拆分为子任务链结构,并收集每个子任务的候选制造服务以及用户消费历史数据;根据Deep Q-Learning框架的状态选择优质候选服务;使用NSGA-III算法根据子任务链结构挑选优质服务组成服务组合方案;BP+GA算法挖掘用户消费历史数据中隐含的用户偏好,根据用户偏好对高质量服务组合方案进行评分;收集以上步骤产生的状态、奖励、动作数据,用于训练Deep Q-Learning框架中的全连接神经网络;不断循环以上步骤,直到用户满意为止。该方法以Deep Q-Learning为框架,在其中嵌入NSGA-III算法与BP+GA算法,解决因服务组合的多样性和异质性,而导致单个服务推荐算法的推荐结果难使用户满意的问题,实现个性化服务组合推荐。



CN 117314582 A

1. 一种基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法,其特征在于,具体包括如下步骤:

步骤1,将复杂制造任务拆分为子任务链结构,并为每个子任务匹配候选制造服务数据集,收集用户消费历史数据。

步骤2,初始化Deep Q-Learning框架、NSGA-III算法、BP+GA算法中的参数;

步骤3,根据Deep Q-Learning框架的状态选择优质候选制造服务;

步骤4,NSGA-III算法根据子任务链结构挑选优质制造服务组成服务组合方案,并对服务组合方案进行优化,得到高质量制造服务组合方案;

步骤5,BP+GA算法挖掘用户消费历史数据中隐含的用户偏好,根据用户偏好对高质量制造服务组合方案进行评分,将评分结果作为奖励,同时向用户推荐排名前N的制造服务组合方案,如果用户满意则结束流程,否则执行步骤6;

步骤6,根据Deep Q-Learning框架的奖励选择动作,修改Deep Q-Learning框架的状态,转入步骤2。

2. 根据权利要求1所述的一种基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法,其特征在于,复杂制造任务所拆分的子任务链结构可以由串联、并联、选择、循环结构中的一种或者多种组成。

3. 根据权利要求1所述的一种基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法,其特征在于,通过状态与对应候选制造服务属性进行点乘,根据点乘结果排序挑选优质候选制造服务;通过BP+GA算法评估服务组合方案,评估结果作为奖励;动作依据 $\epsilon$ -贪婪策略与评估网络选择。

4. 根据权利要求1所述的一种基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法,其特征在于,NSGA-III算法会根据子任务结构挑选优质候选服务组成服务组合方案。

5. 根据权利要求1所述的一种基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法,其特征在于,BP+GA算法通过优化BP网络的权重和偏差,拟合用户消费历史数据以挖掘隐含的用户偏好。

6. 根据权利要求3所述的一种基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法,其特征在于,每运行200步,将目标网络参数赋予网络结构相同的评估网络;每运行10步,会进行网络训练,目标网络损失为目标网络预测的奖励与环境反馈的真实奖励的差,通过梯度下降法减低损失。

7. 根据权利要求4所述的一种基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法,其特征在于,NSGA-III采用的编码方式为整数编码方式,每个整数都代表一个候选制造服务,编码长度为子任务链长度。

8. 根据权利要求4所述的一种基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法,其特征在于,NSGA-III算法的优化目标与约束为:

$$\begin{cases} \min f_1(x) = \sum_{i=1}^{ST_s} \sum_{j=1}^{SP_i} Q_c x_{ij} \\ \min f_1(x) = \sum_{i=1}^{ST_s} \sum_{j=1}^{SP_i} Q_t x_{ij} \\ \max f_2(x) = \sum_{i=1}^{ST_s} \sum_{j=1}^{SP_i} Q_{Rel} x_{ij} \\ \max f_3(x) = \sum_{i=1}^{ST_s} \sum_{j=1}^{SP_i} Q_{CA} x_{ij} \end{cases} \quad (2)$$

$$\sum_{j=1}^n x_{ij} = 1, x_{ij} \in \{0,1\}, 1 \leq i \leq m, 1 \leq j \leq n \quad (3)$$

式中,  $Q_t, Q_c, Q_{Rel}, Q_{CA}$  分别代表制造服务链的时间、成本、可靠性协作能力,  $ST_s$  代表制造服务链中子任务的数量,  $SP_i$  代表  $s_i$  服务的优质候选服务,  $x_{ij}$  是二进制变量, 当制造服务被选中时为1, 否则为0。

9. 根据权利要求5所述的一种基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法, 其特征在于, BP+GA算法的运行过程分为两个主要阶段: 首先, 在BP阶段, 通过反向传播神经网络来训练模型, 调整神经网络的权重和偏差以适应训练数据。然后, 在GA阶段, 使用遗传算法来搜索更优的权重和偏差配置, 通过不断迭代选择、交叉和变异操作, 生成不同的个体并评估其性能, 最终获得最佳的神经网络配置。

10. 根据权利要求6所述的一种基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法, 其特征在于, 目标网络的训练数据采取采用随机抽样的方式, 从记忆库中选择小批量的经验数据。这有助于减少数据之间的相关性, 降低训练的方差, 从而提高训练的稳定性。

## 基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法

### 技术领域

[0001] 本申请涉及云制造服务组合领域与个性化推荐领域,尤其是涉及一种基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法。

### 技术背景

[0002] 随着云计算、物联网和数据分析等先进技术的不断发展,制造过程逐渐数字化、虚拟化,这导致了云制造平台的出现和逐渐成熟。不同制造企业都可将其制造资源打包为制造服务,并将其交付给平台,导致平台上服务种类繁多,属性各异。平台帮助用户获得合适的制造服务,以完成用户需求。随着平台上用户数量以及用户定制化需求的增加,制造需求将面临众多可选服务,这增加了选择的复杂性。

[0003] 个性化推荐算法已在一些大型互联网平台上得到应用,如:淘宝、京东、亚马逊。这些平台基于历史购买数据与个人信息分析用户偏好,根据偏好为用户推荐个性化产品。这些推荐算法大多为单个服务设计。由于制造过程的复杂性,往往需要制造服务组合才能完成制造过程。由于服务组合的多样性和异质性,现应用在淘宝等平台的个性化推荐算法难以应用于服务组合问题中。当制造服务被组合调用时,服务间具有隐藏的协作关系。然而,在大多数个性化制造服务推荐方法中,没有考虑服务的协作关系,这可能造成制造任务执行失败。例如,某些制造服务常一起被成功调用,推荐这些具有紧密协作关系的制造服务,可以减小制造任务执行失败的概率。

### 发明内容

[0004] 本发明的目的在于针对上述问题,提供一种基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法。

[0005] 实现本发明目的的技术解决方案为:基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法,所述方法包括以下步骤:

[0006] 步骤1,将复杂制造任务拆分为子任务链结构,并为每个子任务匹配候选服务数据集,收集用户消费历史数据。

[0007] 步骤2,初始化Deep Q-Learning框架、NSGA-III算法、BP+GA算法中的参数;

[0008] 步骤3,根据Deep Q-Learning框架的状态选择优质候选制造服务;

[0009] 步骤4,NSGA-III算法根据子任务链结构挑选优质制造服务组成服务组合方案,并对服务组合方案进行优化,得到高质量服务组合方案;

[0010] 步骤5,BP+GA算法挖掘用户消费历史数据中隐含的用户偏好,根据用户偏好对高质量服务组合方案进行评分,将评分结果作为奖励,同时向用户推荐排名前N的服务组合方案,如果用户满意则结束流程,否则执行步骤6;

[0011] 步骤6,根据Deep Q-Learning框架的奖励选择动作,修改Deep Q-Learning框架的状态,转入步骤2。

[0012] 进一步地,步骤2中所述的Deep Q-Learning框架包括:智能代理、状态、动作、奖励、深度神经网络。

[0013] 智能代理与环境互动,通过观察环境的状态,选择动作来最大化累积的奖励信号;状态是智能代理观察到的环境信息;动作是代理程序执行的操作;奖励是一个数值信号,表示代理程序的行为好坏;深度神经网络被用来近似代理程序的策略函数,网络结构通过式(1)确定。

$$k = \sqrt{n+m} + t \quad (1)$$

式中,k是隐藏层中节点的数量,n表示输入层中的节点数量,m表示输出层中节点数量,t是[1,10]之间的常数。

[0014] 进一步地,步骤4中所述的NSGA-III算法以多个候选服务属性为优化目标(本申请中以时间、成本、可靠性、协作能力为优化目标),采用整数编码方式,每个整数代表一个候选制造服务,编码长度为子任务链长度。

[0015] NSGA-III算法通过随机生成一组初始个体,创建一个种群。然后,利用非支配排序和拥挤度距离计算,将个体划分为不同的非支配前沿,并选择最优的个体以构建初始解集。接下来,采用进化操作如交叉和变异来生成新的个体,并结合非支配排序和拥挤度距离来选择下一代的解集。这个过程不断迭代,直到达到停止条件,最终产生一组平衡多目标解决方案。

[0016] 进一步地,步骤5所述的BP+GA算法通过式(1)确定BP网络结构,使用GA算法优化BP网络权重与偏差。

[0017] BP+GA算法运行过程如下:首先,通过对BP网络进行初始化,使用用户消费历史数据来训练神经网络模型,得到一个初始的权重和偏差配置。接下来,GA算法被用于进一步优化神经网络的权重和偏差。GA通过生成一组个体(即一组不同的权重和偏差配置)来构建一个种群。然后,对这些个体进行评估,以确定每个个体的适应度。接着,GA通过选择、交叉和变异等操作来演化种群,生成下一代个体。这个过程重复进行多代,直到达到满意的性能优化程度为止。

[0018] 本发明与现有技术相比,其显著优点为:

[0019] 基于Deep Q-Learning框架进行制造服务组合与个性化推荐,可以适应云制造环境中制造服务的动态变化。

[0020] 基于Deep Q-Learning的状态挑选优质候选服务,在保证解质量的同时,减轻NSGA-III算法计算量。

[0021] 使用GA算法优化BP网络权重与偏差,快速准确地挖掘用户偏好。

## 附图说明

[0022] 本申请上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解,其中:

[0023] 图1为本申请实施例所提供的基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法的算法流程图。

[0024] 图2为本申请实施例所提供的基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法的数据传递方式。

[0025] 图3为本申请实施例所提供的NSGA-III算法流程图

[0026] 图4为本申请实施例所提供的为BP+GA算法流程图  
具体实施方法

[0027] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处描述的具体实施例仅用以解释本申请,并不用于限定本申请。

[0028] 在一个实施例中,结合图1、图2,提供基于Deep Q-Learning框架的三阶段个性化制造服务组合推荐方法,所述方法包括以下步骤:

[0029] 步骤1,将用户提交的复杂制造任务拆分为子任务链结构,并为每个子任务匹配候选服务数据集,同时收集用户消费历史数据;初始化评估网络与目标网络,网络结构由式(1)确定。

[0030] 步骤2,通过Deep Q-Learning框架状态与对应候选服务属性进行点乘运行,对结果进行降序排列,挑选前若干个候选服务为优质候选服务。

[0031] 步骤3,NSGA-III算法对优质候选制造服务按照子任务链结构进行服务组合,通过循环优化,得到高质量服务组合方案。

[0032] 步骤4,BP+GA算法挖掘用户消费历史数据中的隐含的用户偏好,通过用户偏好对服务组合方案进行评分,将方案评分定义为Deep Q-Learning框架的奖励。

[0033] 步骤5,通过一个随机数判断是否使用 $\epsilon$ -贪婪策略选择动作;若是,则通过评估网络选择一个奖励值最大的动作;若否,则随机选择一个动作。

[0034] 步骤6,执行动作,修改状态,得到新状态。

[0035] 步骤7,将状态、动作、奖励、新状态保存到记忆库中。

[0036] 步骤8,根据循环次数判断是否要进行目标网络训练、网络间参数转移;若循环次数是10的整数倍,则抽取记忆库中数据训练目标网络;若循环次数是200的整数倍,则将目标网络参数赋予评估网络;若以上条件均不满足,则转入第2步。

[0037] 进一步地,在步骤1中复杂制造任务所拆分的子任务链结构可以由串联、并联、选择、循环结构中的一种或者多种组成。

[0038] 评估网络与目标网络为全连接神经网络,且评估网络与目标网络的结构是完全相同的。

[0039] 进一步地,在步骤3中NSGA-III算法会对候选服务的某些属性进行优化,本申请的优化目标是使时间、成本、可靠性、协作能力最优。优化目标函数为:

$$\begin{cases} \min f_1(x) = \sum_{i=1}^{ST_t} \sum_{j=1}^{SP_i} Q_c x_{ij} \\ \min f_1(x) = \sum_{i=1}^{ST_s} \sum_{j=1}^{SP_i} Q_t x_{ij} \\ \max f_2(x) = \sum_{i=1}^{ST_s} \sum_{j=1}^{SP_i} Q_{Rel} x_{ij} \\ \max f_3(x) = \sum_{i=1}^{ST_s} \sum_{j=1}^{SP_i} Q_{CA} x_{ij} \end{cases} \quad (2)$$

约束条件为:

$$\sum_{j=1}^n x_{ij} = 1, x_{ij} \in \{0,1\}, 1 \leq i \leq m, 1 \leq j \leq n \quad (3)$$

式中, $Q_t, Q_c, Q_{Rel}, Q_{CA}$ 分别代表制造服务链的时间、成本、可靠性协作能力, $ST_s$ 代表

制造服务链中子任务的数量,  $SP_i$  代表  $s_i$  制造服务的优质候选服务,  $x_{ij}$  是二进制变量, 当服务被选中时为1, 否则为0。

[0040] NSGA-III采用的编码方式为整数编码方式, 每个整数都代表一个候选服务, 编码长度为子任务链长度。

[0041] NSGA-III的运行流程如图3所示; 首先初始化种群与解空间, 并对个体进行适应度评估, 然后, 在每一代中, 进行交叉和变异操作以产生新个体, 接着通过个体适应度评估选择最优个体, 接着, 通过交叉和变异操作生成新的个体, 并结合非支配排序和拥挤度距离选择下一代的解集, 迭代此过程, 直到达到停止条件。

[0042] NSGA-III算法的个体评估规则主要包括两个关键步骤: 首先, 通过非支配排序将个体划分为不同的非支配前沿, 以识别出Pareto前沿的解决方案以及次优解前沿; 接着, 通过计算拥挤度距离来度量每个前沿中个体的分布密度, 然后选择具有较高拥挤度距离的个体, 以确保解集保持多样性和均衡性, 这样能有效地维护优秀解的群体。

[0043] 进一步地, 在步骤4中BP+GA算法通过优化BP网络的权重和偏差, 拟合用户消费历史数据以挖掘隐含的用户偏好。

[0044] BP+GA算法中BP网络的权重和偏差的优化过程为: 首先, 通过BP网络阶段, 该算法通过反复迭代, 根据已知数据来微调神经网络的权重和偏差, 以最小化预测误差为优化目标。然后, GA被引入, 它以全局搜索的方式探索更广泛的参数空间, 生成不同的权重和偏差组合。GA通过评估这些组合的性能(通常使用神经网络在验证数据上的表现), 选择最佳组合, 并应用遗传算子(如交叉和变异)来生成下一代组合。这个过程不断迭代, 直到达到性能收敛或预设的停止条件。

[0045] BP+GA算法的运行过程分为两个主要阶段: 首先, 在BP阶段, 通过反向传播神经网络来训练模型, 调整神经网络的权重和偏差以适应训练数据。然后, 在GA阶段, 使用遗传算法来搜索更优的权重和偏差配置, 通过不断迭代选择、交叉和变异操作, 生成不同的个体并评估其性能, 最终获得最佳的神经网络配置。

[0046] BP阶段: 前向传播, 使用初始的权重和偏差配置, 将训练数据逐层传递到神经网络, 计算神经网络的输出; 计算损失, 计算神经网络输出与实际目标值之间的误差, 通常使用损失函数衡量; 反向传播误差, 通过反向传播算法, 将误差信号从输出层逐层传播回神经网络, 计算每个权重和偏差对损失的贡献度; 权重更新, 使用梯度下降或其他优化算法, 根据误差信号来更新神经网络的权重和偏差, 以减小损失函数的值。这个过程迭代多次, 直到损失函数收敛或达到预设的停止条件。

[0047] GA阶段: 初始化种群, 创建一个包含多个个体的种群, 每个个体代表一个神经网络的不同权重和偏差配置; 评估适应度, 使用神经网络的数据拟合准确率来评估每个个体的适应度; 选择操作, 根据适应度选择一部分个体作为父代, 通常选择适应度较高的个体, 以提高下一代的性能; 交叉操作: 对选定的父代个体进行交叉操作, 生成新的个体, 通过组合不同个体的权重和偏差配置, 以增加多样性; 变异操作: 对新生成的个体进行变异操作, 引入随机性, 以进一步增加多样性和探索性; 进化: 上述选择、交叉和变异操作迭代进行多代, 直到达到停止条件, 通常是达到一定的迭代次数或性能收敛。

[0048] BP与GA结合阶段: BP+GA算法将GA阶段生成的最佳个体的权重和偏差配置与BP阶段训练的神经网络模型相结合, 以获得经过优化的神经网络模型。这个结合过程可以提高

神经网络的性能和泛化能力,帮助克服局部最优问题,从而更好地优化神经网络的权重和偏差。

[0049] 进一步地,在步骤5中评估网络的输入为状态值,通过神经网络中的参数传递,最后输出奖励值最大的动作。

[0050] 进一步地,在步骤7中的记忆库是深度强化学习的一种关键技术,用于存储代理程序在与环境互动中获得的经验,经验包括状态、动作、奖励、新状态。

[0051] 进一步地,在步骤8中神经网络损失为目标网络预测的奖励与环境反馈的真实奖励的差,通过梯度下降法减低损失。

[0052] 由记忆库抽取记忆需满足以下规则:随机抽样,采用随机抽样的方式,从记忆库中选择小批量的经验数据。这有助于减少数据之间的相关性,降低训练的方差,从而提高训练的稳定性;经验多样性:记忆库中应包含多样性的经验,以便代理程序能够学习应对各种情况。这意味着在记忆库中应包含来自不同时间步、不同状态和不同行动的经验数据。



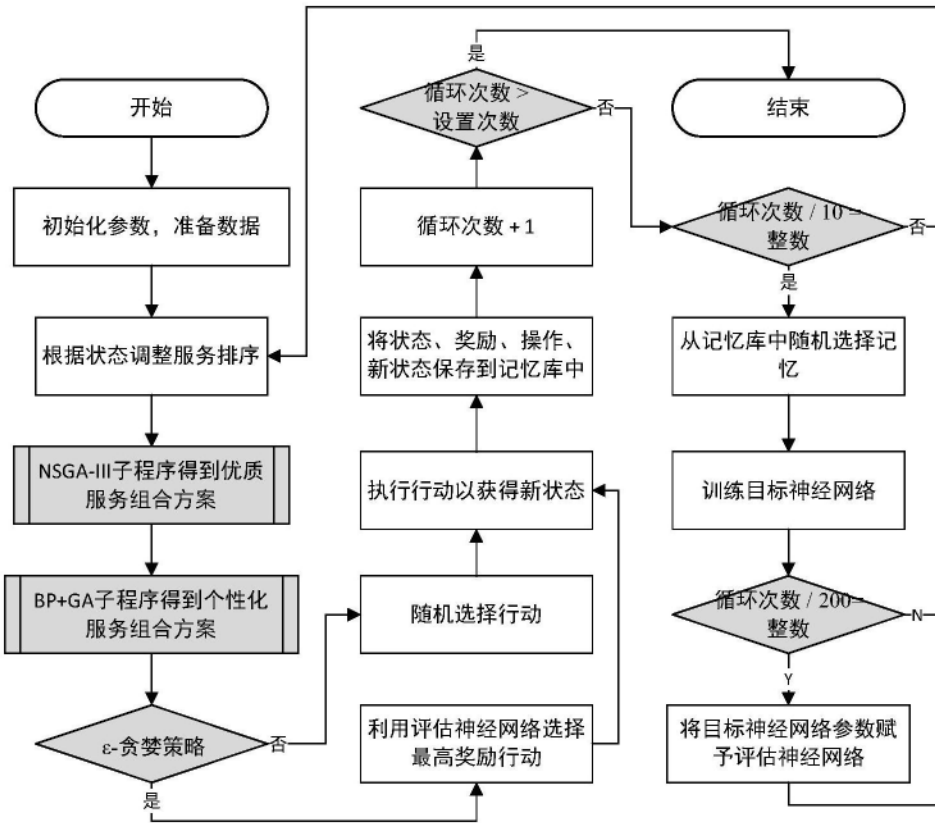


图1

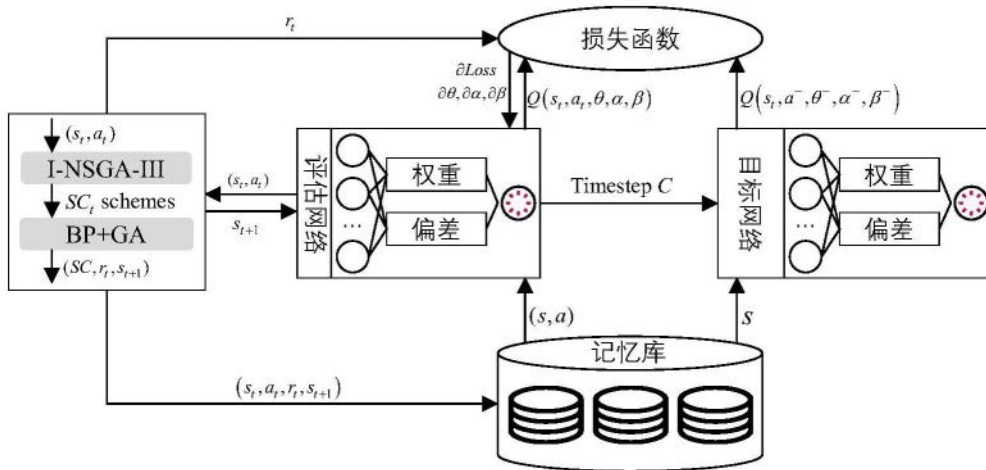


图2

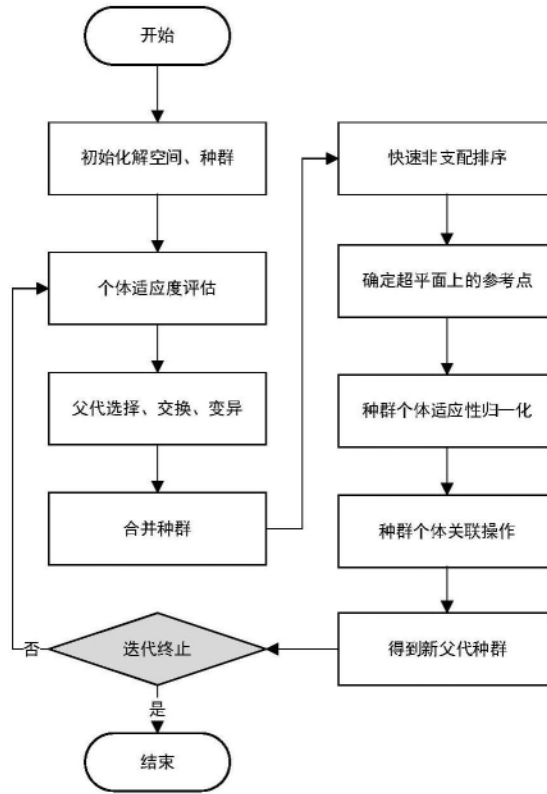


图3

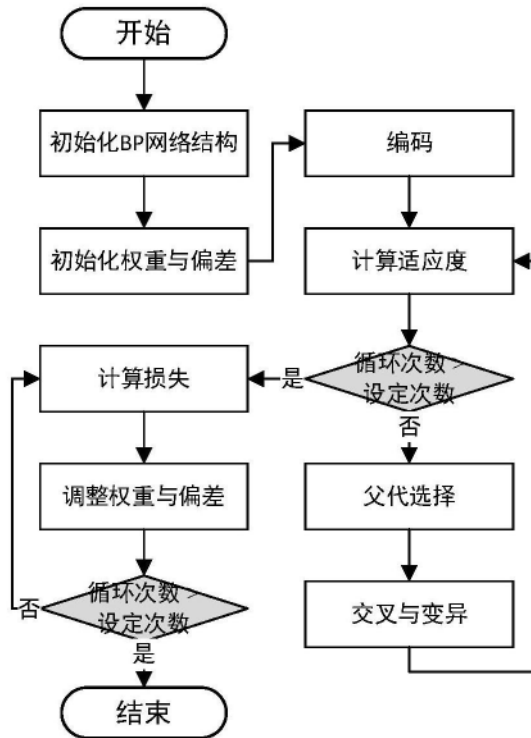


图4