



(12) 发明专利申请

(10) 申请公布号 CN 101751437 A

(43) 申请公布日 2010.06.23

(21) 申请号 200810240358.6

(22) 申请日 2008.12.17

(71) 申请人 中国科学院自动化研究所

地址 100080 北京市海淀区中关村东路 95 号

(72) 发明人 杨彦武 张文生 李益群 肖宪 刘琰琼 梁玉旋

(74) 专利代理机构 中科专利商标代理有限责任公司 11021

代理人 周国城

(51) Int. Cl.

G06F 17/30(2006.01)

G06N 1/00(2006.01)

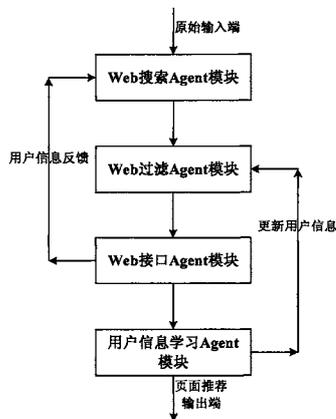
权利要求书 1 页 说明书 4 页 附图 3 页

(54) 发明名称

基于强化学习的网页页面主动式检索系统

(57) 摘要

本发明公开了一种基于强化学习的网页页面(Web)主动式检索系统,该系统包括:Web搜索Agent模块,用于完成基于用户主要兴趣主题... 基于强化学习的网页页面主动式检索系统... 具有自适应能力强,准确度高,使用方便的特点。



1. 一种基于强化学习的网页页面主动式检索系统,其特征在于,该系统包括:

Web 搜索 Agent 模块,用于接收用户的初始请求,并且将用户请求进行分析,利用相关主题分析下载 Web 页面,将结果提交至 Web 过滤 Agent 模块;

Web 过滤 Agent 模块,用于实现对信息搜索 Agent 搜索获取的页面进行内容分析,利用强化学习中的 Q 学习系统对各个 Web 页面进行 Q 学习中值函数计算,并将结果提供 Web 接口 Agent 模块;

Web 接口 Agent 模块用于提供用户推荐 Web 页面,记录用户浏览行为,并将结果提交给用户信息学习 Agent 模块;

用户信息学习 Agent 模块,用于利用强化学习中的 TD 学习算法对用户的兴趣模型进行更新和改进,根据 Web 接口 Agent 模块所完成的对用户浏览页面行为的记录,生成获取到的用户的反馈值,由兴趣度计算和更新模块不断对用户的兴趣进行更新,利用 TD 学习算法对用户的兴趣模式进行计算,最终达到用户信息模型的最佳权重分布。

2. 根据权利要求 1 所述的基于强化学习的网页页面主动式检索系统,其特征在于,该 Web 搜索 Agent 模块由信息搜索、Web 页面分析以及 Web 页面下载这几个功能模块组成,用于实现与用户兴趣主题相关的搜索、网页内容分析和页面的下载功能;由用户先输入原始的请求,根据搜索引擎进行初始页面的获取,并且将页面上的相关链接提取并存放到一个缓冲区,页面下载模块将根据链接 URL 地址访问相应的网页,同时按照主题关键词分类保存。

3. 根据权利要求 1 所述的基于强化学习的网页页面主动式检索系统,其特征在于,该 Web 过滤 Agent 模块接收到 Web 搜索 Agent 模块实施结果之后,利用 Q 学习算法结合用户特征模型对 Web 页面内容进行分析及排序,取出 Q 值排名靠前的 Web 页面进行过滤;Web 过滤 Agent 模块的作用主要包括 Web 页面分析、Web 页面过滤和 Web 页面分类索引,并将结果提供 Web 接口 Agent 模块。

4. 根据权利要求 1 所述的基于强化学习的网页页面主动式检索系统,其特征在于,该 Web 接口 Agent 模块用于实现推荐与学习模块完成的用户模型最相关的网页予以浏览、并且接收反馈值,对浏览行为的分析和统计。

5. 根据权利要求 1 所述的基于强化学习的网页页面主动式检索系统,其特征在于,该用户信息学习 Agent 模块用于根据知识库初始生成该用户的基本信息文件,在将来的搜索以及浏览网页的过程中,不断对用户的兴趣进行更新,以确定用户信息模型。

6. 根据权利要求 5 所述的基于强化学习的网页页面主动式检索系统,其特征在于,该用户信息学习 Agent 模块根据用户反馈值,进一步对用户信息模型进行更新,基于 TD 算法对用户模型进行迭代计算和更新并判断用户模型是否达到收敛。

7. 根据权利要求 5 或 6 所述的基于强化学习的网页页面主动式检索系统,其特征在于,该用户信息学习 Agent 模块将运行结果提交给 Web 过滤 Agent,进入下一步学习。

8. 根据权利要求 5 或 6 所述的基于强化学习的网页页面主动式检索系统,其特征在于,该用户信息学习 Agent 模块在迭代结果达到收敛时,用户信息模型达到最优,并根据用户信息模型,获取推荐 Web 页面。

基于强化学习的网页页面主动式检索系统

技术领域

[0001] 本发明涉及 Web 用户的主动式检索技术领域,尤其涉及一种基于强化学习的网页页面主动式检索系统,用于实现对 Web 用户进行最能体现用户兴趣模式的 Web 页面推荐。

背景技术

[0002] 马尔科夫决策过程包含一个环境状态集 S , 方法行为集合 A , 奖赏函数 R 和状态转移函数 P 。奖赏函数 $R(s, a, s')$ 是在状态 s 的情形下采用动作 a , 环境状态转移到 s' 获得的瞬时奖赏值;记 $P(s, a, s')$ 在状态 s 的情形下采用动作 a 使环境状态转移到 s' 的概率。马尔科夫决策过程的本质是:当前状态向下一状态转移的概率和奖赏值只和当前状态以及当前状态下选择的动作有关,而与以前的历史状态以及历史动作无关。因此在状态转移概率函数 P 和奖赏函数 R 都已经确定的环境模型的知识框架下,动态规划的技术可以用来求解最优策略。然而在现实世界中的大部分情况下,状态转移概率函数 P 和奖赏函数 R 的环境模型却难以确定,强化学习主要是着重研究奖赏函数和状态转移函数未知的情况下,如何学习最优行为策略。

[0003] 强化学习 (reinforcement learning, 又称再励学习, 评价学习) 是机器学习方法的一个重要的分支,在智能控制机器人及分析预测等领域有许多应用。强化学习是对智能系统中从环境到行为映射的学习,以使累积的奖赏(强化信号)函数值最大,强化学习不同于传统机器学习中的监督学习主要表现在教师信号上,强化学习中由环境提供的强化信号是对做出的动作作出一种评价作为奖赏值,而不是告诉直接强化学习系统 (reinforcement learning system) 如何去产生正确的动作。由于外部环境提供的信息较少,强化学习系统必须靠自己获得的经历进行学习。通过这种方式,强化学习系统在行动-评价的环境中获得计算基础,提出改进的行动方案以适应环境。目前的强化学习的学习技术大致可分成两类:一是搜索智能系统的行为空间,从而发现可以做出的最优的行为。典型的技术如遗传算法等搜索技术;二是采用基于统计的技术和动态规划的思想来估计和预测在某一确定环境状态下的价值函数值,从而通过获取的价值函数来确定最优行为。

[0004] 在强化学习需要解决的问题中,由于环境是不确定的,策略 π 指导下的每一次学习所得到的 R_t 有可能是不相同的。因此在 s 状态下的值函数要考虑在不同学习中所有可能的返回函数的数学期望值。实际中经常采用逼近方法进行值函数的估计,一种最主要的方法就是 Monte Carlo 采样方法。将 Monte Carlo 采样方法和动态规划技术结合起来,通过多次试验,用实际获得的奖惩返回值去逼近真实的状态值函数, Monte Carlo 采样方法通常是采用一次学习循环所获得的值函数去逼近实际的值函数,而强化学习方法使用下一状态的值函数(即 Bootstrapping 方法)和当前获得的瞬时奖赏来逼近当前状态值函数。强化学习方法需要多次学习循环才能最终逼近实际的值函数。

[0005] 信息检索 (Information Retrieval), 通常指的是基于文本的信息检索,包括信息的存储、组织、表现、查询、存取等各个方面,其核心为文本信息的索引和检索。从历史发展进度来看,信息检索经历了人类手工检索,计算机自动化检索,网络智能化检索等多个发展

阶段。目前,信息检索已经发展到网络化和智能化的阶段。信息检索的对象也从封闭、比较稳定一致、由独立的数据库集中管理信息内容扩展到开放、动态、快速、分布广泛、管理松散复杂的 Web 页面内容;原来的使用信息检索的用户为情报专业人员,现在的信息检索包括商务人员、管理人员、教师、学生、各专业人士等在内的普通大众,他们对信息检索从结果到方式提出了更高、更多样化的要求。适应网络化、智能化以及 Web 个性化的需要是目前信息检索技术发展的新趋势。现实世界中,目前存在较多基于统计方法的对于网络个性化的方法。但是这种方法自适应能力较差,并且不具备学习能力。然而强化学习的特点可以改进目前的这种基于统计方法的网络个性化分析方法。

发明内容

[0006] (一) 要解决的技术问题

[0007] 有鉴于此,本发明的主要目的在于提供一种基于强化学习的 Web 主动式检索系统,以协助用户更方便的浏览 Web,以及更准确的找到用户所需要的目标页面。

[0008] (二) 技术方案

[0009] 为达到上述目的,本发明提供了一种基于强化学习的 Web 主动式检索系统,该系统包括:

[0010] 1) Web 搜索 Agent 模块,由信息搜索、Web 页面分析以及 Web 页面下载这几个功能块组成,主要完成利用与用户兴趣主题相关的搜索、网页内容分析和页面的下载功能。首先,由用户先输入原始的请求,根据搜索引擎进行初始页面的获取,并且将页面上的相关链接提取并存放到一个缓冲区,页面下载模块将根据链接(URL 地址)访问相应的网页,同时按照主题关键词分类保存。

[0011] 2) Web 过滤 Agent 模块,由页面分析及页面过滤两个功能块组成,主要完成对信息搜索 Agent 搜索获取的页面进行内容分析,利用强化学习中的 Q 学习系统对各个 Web 页面进行 Q 学习中值函数计算,此时,Web 页面的对于关键词的 TFIDF 值作为 Q 学习中的立即奖赏值,可以计算出各个 Web 页面所对应的 Q 值,对 Web 页面的 Q 值排序和过滤,取出 Q 值排名较靠前的 Web 页面;Web 过滤 Agent 模块的作用主要包括 Web 页面分析、Web 页面过滤和 Web 页面分类索引,并将结果提供 Web 接口 Agent 模块。

[0012] 3) Web 接口 Agent 模块,由页面推荐模块、页面显示模块和浏览行为统计模块组成,完成推荐与学习模块完成的用户模型最相关的网页予以浏览、并且接收反馈值,对浏览行为的分析和统计等功能。

[0013] 4) 用户信息学习 Agent 模块,由初始化、兴趣度修正计算和兴趣度更新等主要功能块组成,首先需要用户的注册,根据知识库的信息自动生成该用户的初始兴趣文件(Profiles),在搜索浏览网页过程中,利用强化学习中的 TD 学习算法可以对用户的兴趣模型进行更新和改进,根据 Web 接口 Agent 模块所完成的对用户浏览页面行为的记录而生成获取到的用户的反馈值,由兴趣度计算和更新模块不断对用户的兴趣进行更新,利用 TD 学习算法对用户的兴趣模式进行计算,最终达到用户信息模型的最佳权重分布。

[0014] (三) 有益效果

[0015] 从上述技术方案可以看出,本发明具有以下有益效果:

[0016] 1、本发明提供的这种 Web 主动式检索系统,具有准确度高,使用方便的特点,能够

实现对 Web 用户进行个性化 Web 页面推荐。

[0017] 2、本发明提供的这种 Web 主动式检索系统,利用强化学习的 Q 学习的系统,对由 Web 搜索 Agent 的页面进行过滤,可以更好的考虑到不止是当前推荐的页面,而且考虑到 Web 搜索 Agent 推荐页面的超链接的网页信息,充分利用了 Web 页面结构来进行过滤系统的优化。

[0018] 3、本发明在进行用户信息学习中,利用了强化学习中的 TD 学习多样化模型进行学习和更新,使得用户模型可以越来越接近用户的最佳代表模型,将模型应用到 Web 过滤 Agent 就可以提取出更能符合用户兴趣模型的 Web 页面予以推荐。

附图说明

[0019] 图 1 是本发明提供的 Web 主动式检索系统的整体逻辑框图;

[0020] 图 2 是本发明提供的 Web 主动式检索系统中对 Web 页面过滤的结构示意图;

[0021] 图 3 是本发明提供的 Web 主动式检索系统的中用户信息学习和更新的结构示意图。

具体实施方式

[0022] 为使本发明的目的、技术方案和优点更加清楚明白,以下结合具体实施例,并参照附图,对本发明进一步详细说明。

[0023] 如图 1 所示,图 1 是本发明提供的 Web 主动式检索系统的整体逻辑框图,包括一个原始输入端和一个页面推荐输出端。其中,原始输入端用于向网页页面 (Web) 搜索 Agent 模块发送原始的用户请求,Web 搜索 Agent 模块接收用户请求后利用搜索引擎进行搜索,搜索完毕后,Web 过滤 Agent 模块对其中的页面进行学习和排序,提取出页面通过 Web 接口 Agent 模块推荐给用户;用户在浏览之后,利用 Web 接口 Agent 模块进行记录,获取用户的浏览信息,进行分析和统计,将反馈值提交给用户信息学习 Agent 模块,用户信息学习 Agent 模块将对其进行更新和优化。

[0024] 如图 2 所示,图 2 是本发明提供的 Web 主动式检索系统中的 Web 过滤 Agent 模块的结构示意图。该 Web 过滤 Agent 模块包括相关值估计模块、Q 学习模块和 Q 值排序模块组成。在 Web 搜索 Agent 模块搜索完毕之后,将存贮的页面取出,进行页面内容的分析,根据 TFIDF 技术,可以将 Web 页面用向量模式表示,利用余弦计算距离的方法计算用户与 Web 页面的相似度,对于一个给定的 i (第 i 个页面文件) 我们需要对其中的每个关键词进行学习,对应的每个页面的 TFIDF 值就作为到达该页面时所获的立即奖赏值,当前页面的超链接所对应的下一层页面的奖赏之可作为下一个状态对应点的奖赏值,在求奖赏值总合时需要乘上一个折扣值,这里,我们采取了基于图的深度优先搜索算法实现计算最大的路径的奖赏值的折扣和,并将其赋值给 $Q[i]$, $Q[i]$ 就记录为当前第 i 个页面 Q 值,计算出所有的页面的 Q 值之后对所有的网页对应的 Q 值进行的。将排名靠前的前 K 个页面进行推荐给用户。

[0025] 如图 3 所示,图 3 是本发明提供的 Web 主动式检索系统中用户信息学习 Agent 模块的结构示意图。该用户信息学习 Agent 模块主要包括计算用户模型更新权重模块和 TD 学习更新模块,在 Web 接口 Agent 获取了用户的浏览 Web 信息之后,Web 接口 Agent 会提供一

个用户的反馈值给用户信息学习 Agent 模块以供用户信息学习模块进行学习和更新,反馈值主要是由显式的和隐式的反馈值组成,其中显式的反馈主要是由用户的评分获取,而隐式的反馈值主要决定于四大要素:bookmarking(bm), reading time(rt), scrolling(sc), following up the hyperlinks in the filtered documents(fl)。利用反馈值可以对用户的权重进行更新:

$$[0026] \quad W_{p,k} \leftarrow W_{p,k} + \beta r_{i,k}$$

[0027] β 为用户学习的速度。 $W_{p,k}$ 表示用户向量的第 k 维的值。

[0028] 利用向量的项权值(归一化处理之后)的变化来近似的衡量未来的奖赏值。当用户模型权重变化小于经过预先测试得到的阈值时表示关键词的选择代表用户的兴趣已经趋于最优,当这个变化之为正时,说明当前用户模型向量将会获得更好的反馈值。

$$[0029] \quad W_{pk,t} \leftarrow W_{pk,t-1} + [R_t + \gamma \Delta v_{p,t}]$$

$$[0030] \quad R_t = \frac{1}{K} \sum_{i=1}^K r_i \Delta v_{p,i} = \frac{1}{K} \sum_{i=1}^K \sum_{j=1}^n (w_{pj,i} - w_{pj,i-1})$$

[0031] 如果 W_{pk} 在推荐的页面中出现,则增加这些出现的关键词权重,如果变化值为负值,则这些关键词的权值的变化较小。处理完这 K 个页面的内容之后,Web 主动式检索系统会进行到下一个检索点,学习将继续进行,一直到能够达到近似代表最优用户模型。

[0032] 本发明提供的 Web 主动式检索系统,具有学习能力强、使用方便、可靠性高的特点,能够方便的使用该系统协助用户进行主动式检索和推荐 Web 页面。

[0033] 以上所述的具体实施例,对本发明的目的、技术方案和有益效果进行了进一步详细说明,所应理解的是,以上所述仅为本发明的具体实施例而已,并不用于限制本发明,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

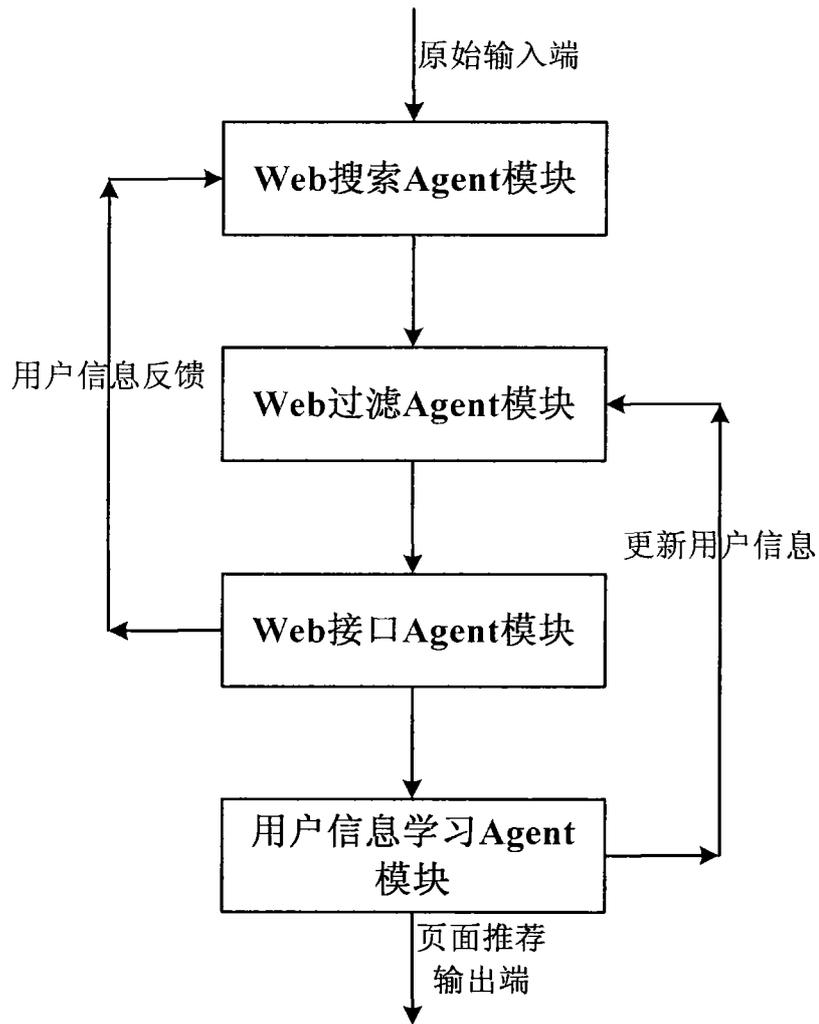


图 1

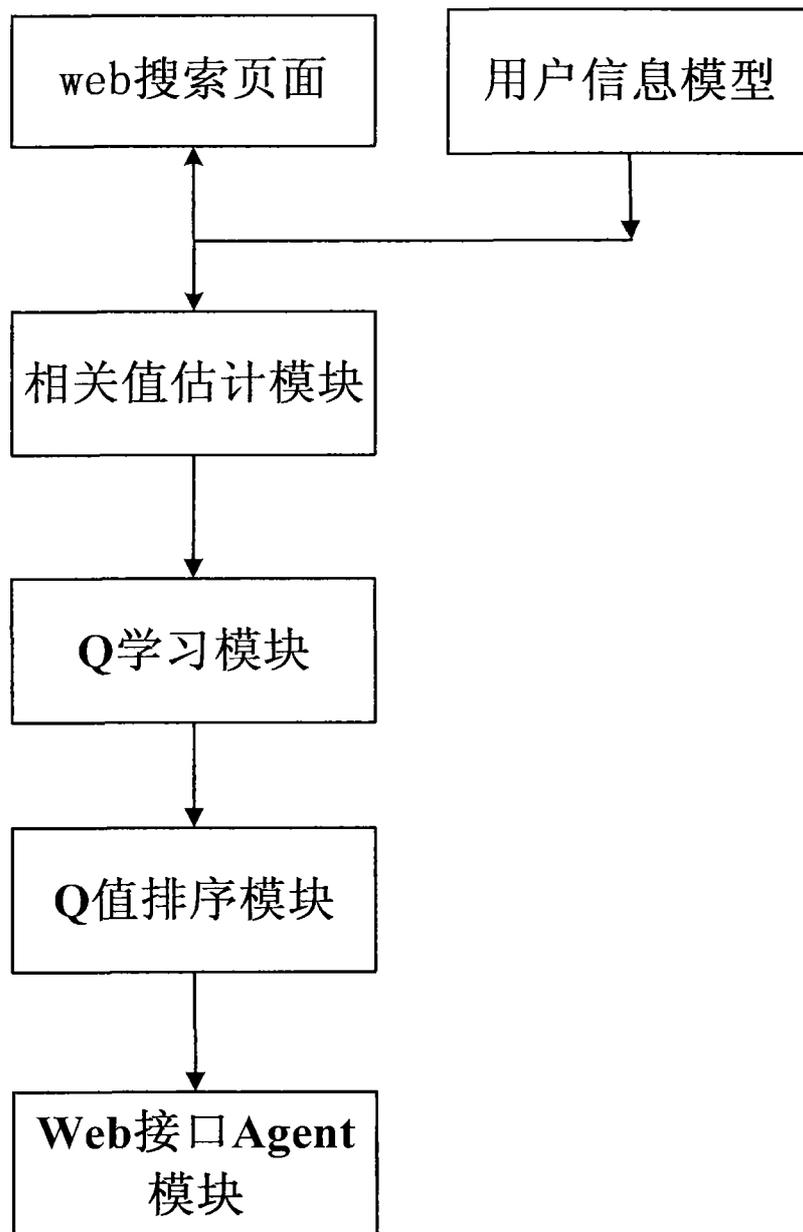


图 2

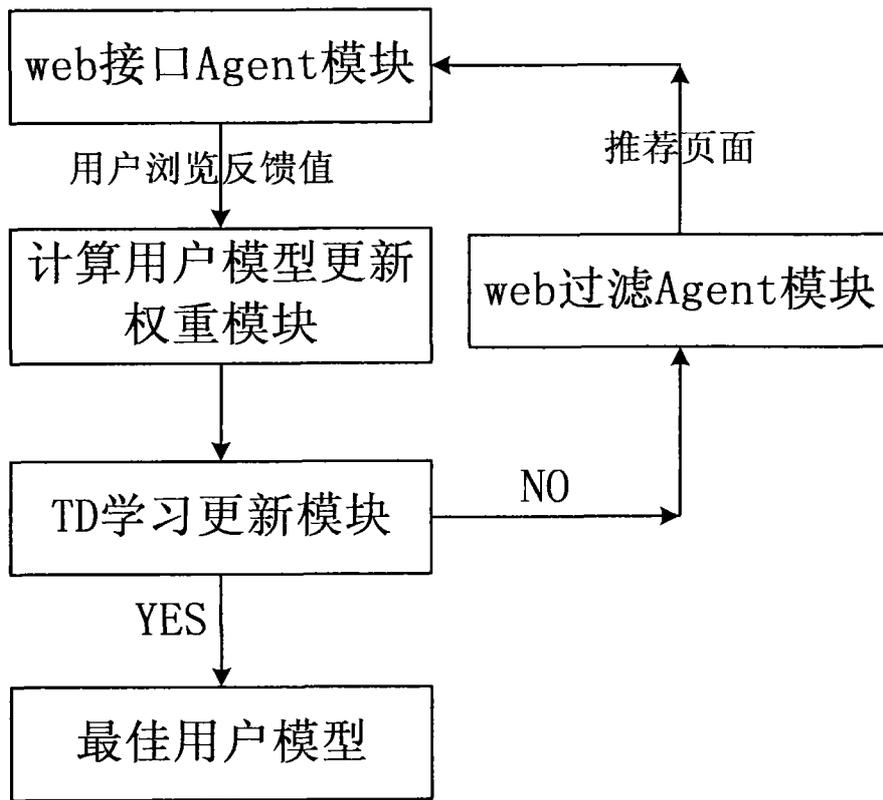


图 3