

项目负责人签字：\_\_\_\_\_ 日期 年 月 日

# 说明书

## 工业平台资源优化分配装置及方法

### 技术领域

[0001] 本发明涉及的是一种工业海量信息处理领域的技术，具体是一种工业平台资源优化分配装置及方法。

### 背景技术

[0002] 随着信息化的发展，工业领域系统的规模越来越大。例如，在大规模分布式任务或系统中(例如覆盖多地区的众包任务、互联网中推荐系统的内容分发任务等)，基于智能算法的机器人(或代理程序等)需要完成各自的任務。但是，当数据量巨大，或者机器人因为某些原因不能向服务器公开自己所有信息的情况时，控制端服务器无法存储所有机器人的信息，并且服务器无法同时向所有的机器人分配所需的资源和任务。

### 发明内容

[0003] 本发明针对现有技术存在的上述不足，提出一种工业平台资源优化分配装置及方法，通过对机器人诉求和优化目标进行建模，本发明推荐给机器人其可以分配到的资源，并通过机器人的反馈获取资源分配的合理性，打破服务器端与机器人之间信息不对称的僵局。

[0004] 本发明是通过以下技术方案实现的：

[0005] 本发明涉及一种工业平台资源优化分配装置，包括：内容分发系统和资源库，其中：内容分发系统生成机器人的资源预测请求并输出至资源库，根据资源库的反馈进行最优资源配置，实现机器人服务流程的同时基于新增的数据更新内容分发系统中诉求预测单元中的神经网络模型，资源库接收内容分发系统发出的资源预测请求，预测潜在可分配的最优资源配置，接收内容分发系统的资源调度单元的资源申请，基于资源申请分配资源。

[0006] 所述的内容分发系统包括：交互单元、诉求预测单元、特征存储单元、资源调度单元和网络训练单元，其中：交互单元接收机器人的资源请求，将机器人 ID 和预算发送给诉求预测单元；诉求预测单元将机器人 ID 发送给特征存储单元；特征存储单元将该机器人特征发送给诉求预测单元；诉求预测单元中的神经网络基于机器人特征预测出机器人的诉求，将诉求与预算发送给资源库；诉求预测单元根据来自资源库的资源预测结果发送给交互单元，交互单元询问机器人是否采纳；并在机器人采纳该资源调度结果时，将机器人授权的资源调度结果发送给资源调度单元；资源调度单元发送资源申请请求给资源库；资源调度单元将资源发送给机器人；该轮交互结束，交互单元将最新一轮的交互数据发送给特征存储单元。

[0007] 所述的神经网络模型，通过以下方式进行训练：网络训练单元向特征存储单元发送数据

请求；特征存储单元将训练数据发送给网络训练单元；网络训练单元训练神经网络模型，并更新诉求预测单元中的神经网络模型。

## 技术效果

[0008] 本发明整体解决了现有技术中由于机器人通信或者表达能力限制或者由于资源分发系统存储处理能力和计算能力限制下，机器人诉求难以明确表达，进而难以得到个性化满足，从而导致系统资源分配效率低下的缺点。

[0009] 与现有技术相比，本发明通过对机器人诉求和优化目标进行建模，我们推荐给机器人其可以分配到的资源，并通过机器人的反馈获取资源分配的合理性。我们设计的系统可以依据机器人的采纳行为，收集机器人对不同资源的需求信息，打破服务器端与机器人之间信息不对称的僵局。

## 附图说明

[0010] 图 1 为本发明系统示意图；

[0011] 图 2 为内容分发系统内部结构示意图；

[0012] 图 3 为实施例模型对比实验结果示意图；

[0013] 图中：a)不同比例 Dropout 下的累积期望遗憾，b)不同比例 Dropout 下的累积采纳率；

[0014] 图 4 为实施例诉求相关信息的影响示意图；

[0015] 图中：a)诉求相关信息对累积期望遗憾的影响，b)诉求相关信息对累积采纳率的影响。

## 具体实施方式

[0016] 如图 1 所示，为本实施例涉及一种工业平台信息优化分发装置，包括：内容分发系统和资源库，其中：资源库根据申请资源预算划拨资源返回至内容分发系统，内容分发系统预测机器人的诉求，对其进行最优资源配置并基于新增的数据更新诉求预测单元 2 中的神经网络模型。内容分发系统接收机器人资源申请预算信息，基于机器人的历史数据进行诉求预测，为机器人划拨最优资源配置。

[0017] 如图 2 所示，所述的内容分发系统包括：交互单元 1、诉求预测单元 2、特征存储单元 3、资源调度单元 4、网络训练单元 5，其中：交互单元 1 接收机器人的资源请求，将机器人 ID 和预算发送给诉求预测单元 2；诉求预测单元 2 将机器人 ID 发送给特征存储单元 3；特征存储单元 3 将该机器人特征发送给诉求预测单元 2；诉求预测单元 2 中的神经网络基于机器人特征预测出机器人的诉求，将诉求与预算发送给资源库；资源库计算潜在可分配资源，发送给诉求预测单元 2；诉求预测单元 2 将资源预测结果发送给交互单元 1，交互单元 1 询问机器人是否采纳；并在机器人采纳该资源调度结果，交互单元 1 将机器人授权的资源调度结果发送给资源调度单元 4；资源调度单元 4 发送资源申请请求给资源库；资源库分配相应的资源给资源调度

单元 4；资源调度单元 4 将资源发送给机器人；该轮交互结束，交互单元 1 将最新一轮的交互数据发送给特征存储单元 3。

[0018] 所述的神经网络模型的训练流程中步骤如下：(a)网络训练单元 5 向特征存储单元 3 发送数据请求；(b)特征存储单元将训练数据发送给网络训练单元 5；(c)网络训练单元训练神经网络模型，并更新诉求预测单元 2 中的神经网络模型。

[0019] 本实施例涉及上述工业平台信息优化分发装置，通过以下方式进行工业平台信息优化分发处理：当有机器人发起资源申请请求时，内容分发系统从申请请求中解析出机器人相关信息并生成估计的机器人需求，并将机器人需求以及为该机器人分配的预算等信息发送到资源库，询问可分配的资源；资源库根据内容分发系统提供的需求和预算估计能够获取的资源，并将预计的可分配资源  $v = [v_1, v_2, \dots, v_n]^T$  返回，内容分发系统据此向机器人发送资源申请结果，并根据机器人的采纳反馈信号，通过资源库将基于该需求的真实资源结果分配给机器人。

[0020] 综上，该系统可以依据机器人的采纳行为，收集机器人对不同资源的偏好信息，打破服务器端与机器人之间信息不对称的僵局，更好地配置资源库中的整体资源。

[0021] 所述的可分配的资源是指：在预算等各种约束下，机器人可以获得的资源结果，具体为：  $v = [v_1, v_2, \dots, v_n]^T$ ，其中： $n$ 表示资源的种类数，值 $v_i$ 表示第 $i$ 维资源的量。

[0022] 所述的机器人相关信息，包括：机器人拥有的资源申请预算、机器人对于不同的资源的偏好，即诉求权重向量： $w = [w_1, w_2, \dots, w_n]^T$ ，其中： $w_i$ 表示机器人对 $i$ 维报表的偏好权重。

[0023] 本实施例定义 $\Pi$ 为一种资源配置方法。当机器人的诉求向量为 $w^*$ ，服务器可以基于诉求向量配置其最优的资源。令优化目标为 $w^{*T} \cdot v$ ，即资源结果基于诉求的加权求和。对于诉求为 $w^*$ 的机器人，向其推荐最优资源配置策略 $\Pi_{w^*}$ 即为解最优化问题： $\Pi_{w^*} = \arg \max_{\Pi} w^{*T} \cdot$

$v_{\Pi}$ ，其中：资源申请结果 $v_{\Pi}$ 为资源库基于策略 $\Pi$ 可达的最优解， $w^{*T} \cdot v_{\Pi}$ 为机器人基于策略 $\Pi$ 可以获得的效用。因此，权重向量 $w^*$ 可以帮助内容分发平台在资源库中寻找并获取令机器人最满意的资源申请结果。

[0024] 当进行推荐时，当机器人对预期的结果感到满意，则会采纳该建议。否则，机器人则会跳过该建议。基于上述的观察，本实施例将需要优化的问题建模成上下文老虎机问题，并设计具体的算法程序，具体包括：

[0025] 1)状态：平台能够观察到的与机器人相关的信息。如：机器人特征，可以为该机器人分配的预算，机器人的历史询问和采纳与否等。

[0026] 2)动作：机器人诉求向量。算法程序的动作选择空间为高维连续空间。算法程序需要

根据状态中的约束信息，以及自己选择的动作，即预估的机器人诉求向量，向资源库发送请求，获取资源库预测的资源推荐项。

[0027] 3)回报：本实施例将回报设置为机器人的采纳行为。

[0028] 基于上述的建模，上下文老虎机算法会不断地为来访的机器人进行策略推荐，每轮推荐中：

[0029] 1)算法程序观察到该轮推荐中机器人的状态。

[0030] 2)算法程序基于该状态选择一个诉求向量，将诉求向量与预算等约束信息传给资源库，资源库进行预估的资源结果。算法程序将该结果推荐给机器人，并获得机器人的反馈。

[0031] 3)算法程序存储该轮的观察(机器人状态、诉求向量、资源分配结果、机器人反馈)以作为更新自己智能推荐策略的训练数据。

[0032] 动作价值估计：经典的上下文老虎机算法中，根据所观察到的上下文，算法程序会基于一定的策略对老虎机臂进行拉杆，学习每个可选动作的期望价值。在本实施例的问题中，服务器端选择动作的奖励为机器人的采纳与否，最大化奖励即为推荐给机器人其最有可能采纳的策略。

[0033] 本实施例中动作价值估计过程包括：

[0034] 1)基于能够观察到的信息与动作选择策略进行动作选择。

[0035] 2)建立所选动作、申请资源结果与机器人采纳率之间的联系。

[0036] 本实施例首先刻画机器人信息与动作选择之间的联系，即在服务器端能够观测到的信息下，基于一定动作选择策略得到诉求向量 $w = f(x)$ ，其中：函数 $f$ 为多层感知机，表示环境状态 $x$ 到诉求信息 $w$ 之间的映射关系，函数 $f$ 的输入 $x$ 为环境的特征表达。在本实施例的问题中，网络的输出为 $w$ ，网络的监督信息(即动作的价值)为机器人的采纳行为。直观上说，令 $v$ 为 $w$ 下的最优资源分配结果， $w^T \cdot v$ 的值反映了诉求为 $w$ 的机器人能够在平台获取的效用。因此，机器人的采纳率与 $w^T \cdot v$ 在某种程度上呈正相关。

[0037] 本实施例中采用 $p(\text{采纳}) = \sigma(w^T \cdot v)$ ，表示机器人采纳率与 $w^T \cdot v$ 的关系，其中： $\sigma$ 为sigmoid函数，值域为 $[0, 1]$ ，基于 $w$ 的最优竞价结果 $v$ 也作为模型输入的一部分。

[0038] 基于上述方法，可以采用梯度更新的方式更新网络对动作价值的估计。对于每一轮的梯度更新，本实施例通过损失函数 $L$ 更新模型的参数，具体为： $L =$

$-\frac{1}{N} \sum_{(x, v, y) \in \mathcal{D}} (y \log p(x, v) + (1 - y) \log (1 - p(x, v)))$ ，其中：集合 $\mathcal{D}$ 为该轮更新中大小为 $N$ 的数据集，环境特征 $x$ 与资源申请结果 $v$ 为模型的输入， $p(x, v)$ 为模型预测的采纳率，标签 $y$ 为采纳标签。训练过程中，需要先输入环境特征 $x$ ，得到模型的诉求输出 $w$ ，并根据 $w$ 获得结果 $v$ ，最后得到模型的预估采纳率 $p(x, v)$ 。



[0039] 动作选择算法：本实施例使用汤普森采样进行动作选择，汤普森采样是一种在探索(Exploration)和利用(Exploitation)之间权衡的流行手段。通常来说，汤普森采样需要对模型参数进行贝叶斯处理。在每个步骤中，汤普森采样都会重新采样一组新的模型参数，然后基于该组参数进行动作选择。这可以看作是一种随机检验：更可能的参数会被更频繁地采样到，从而更快地被驳斥或者确认。

[0040] 所述的汤普森采样包括：采样模型的一组新参数；根据采样参数选择预期收益最高的动作；更新模型参数。

[0041] 对神经网络模型进行汤普森采样，需要刻画模型的不确定性，贝叶斯模型提供了基于数学框架来推理模型的不确定性，但通常带有禁止性的计算成本。Dropout 是指在神经网络的训练过程中，按照一定的概率将一部分神经元暂时从神经网络中丢弃。Yarin 等人在

《Dropout as a Bayesian approximation: Representing model uncertainty in deep learning》中提出利 Dropout 作为一种贝叶斯近似的方法，在深度学习中表示模型的不确定性，并证明了具有任意深度的非线性神经网络，当在每个权重层之前应用 Dropout，则在数学上等效于深度概率高斯过程的近似。此外，Dropout 作为神经网络的一种简单且常见的防止过拟合的技巧，由于易于实现、性能高效且有着很好的效果，已经被广泛运用于神经网络的训练中。因此，本实施例在神经网络中使用 Dropout 来进行汤普森采样，这非常简单但有效。

[0042] 在实验中，模型的输入特征为诉求相关特征和与机器人的历史采纳信息。对于诉求相关特征，本实施例将其串联起来作为模型的输入之一。训练过程中，本实施例使用小批量梯度下降来训练网络模型。为了防止模型获得的正负样本比例随训练过程而变化，从而影响模型性能，本实施例将每个训练批次中正负样本的比例设置为 1: 1。模型训练过程中的优化器为 Adam，实验结果如表 1 所示。

[0043] 上下文老虎机的优化目标为  $T$  轮推荐期望遗憾，因此，本实施例可以通过累积期望遗憾

$E \left[ \sum_{t=1}^T r_{t, w_t^*} \right] - E \left[ \sum_{t=1}^T r_{t, w_t} \right]$  和累积采纳率  $\frac{\sum_{t=1}^T r_{t, w_t}}{T}$  来评价模型的表现，其中  $T$  表示实验已经进行了  $T$  轮交互，值  $r_{t, w_t^*}$  表示第  $t$  轮中，基于机器人内心诉求  $w_t^*$  进行推荐的采纳率，值  $r_{t, w_t}$  表示第  $t$  轮中，基于动作选择算法输出的  $w_t$  进行推荐的采纳率。

[0044] 表 1

模型	相对累积期望遗憾	相对累积采纳率
随机诉求推荐（无学习模块）	1.0000	0.7606
无 Dropout	0.7429	0.8239
Dropout 比例 40%（无诉求相关信息）	0.7195	0.8535
Dropout 比例 80%	0.5384	0.9144
Dropout 比例 20%	0.4154	0.9725
Dropout 比例 60%	0.4078	0.9817
Dropout 比例 40%	<b>0.3881</b>	<b>1.0000</b>

[0045] 对比实验结果：在仿真实验中，本实施例验证上下文老虎机算法的有效性。对比实验中，本实施例对比了模型在不进行 Dropout 或者在不同比例下进行 Dropout 时的效果。本实施例还引入了不应用任何诉求估计算法的随机诉求推荐策略作为弱基准。每组实验中，算法程序会与环境进行一共 2000 轮的交互，并间隔一定轮数记录当前的累积期望遗憾和累积采纳率，交互结束后的实验结果显示在表 1 中。从结果中，本实施例发现随机诉求推荐系统会导致评价指标的大幅下降，这表明在推荐策略时必须考虑机器人的诉求。不同 Dropout 比例的累积期望遗憾和累积采纳率曲线显示在图 3 中，由于本实施例发现在实验中不同算法会收敛到不同局部最优解，使得期望遗憾在模型收敛后会按一定斜率近似线性增长，为了更好地了解模型收敛后的表现差别，本实施例通过  $y=\log(x+1)$  对累积期望遗憾进行了预处理，并对实验结果进行了归一化以绘制曲线。通过图 3 中的趋势，并分析实验过程中的实时累积期望遗憾和累积采纳率，本实施例发现在图 3 所示的训练过程中，所有模型的累积期望遗憾增量在逐渐减小并收敛。在图 3 所示的训练过程中，所有模型的累积采纳率在逐渐增大并收敛。上述观察说明不同模型收敛到了不同的局部最优解，但是它们都能一定程度地学习到机器人的诉求，并提升推荐系统的表现性能。例如，在表 1 中，即使是没有运用 Dropout 的模型，与随机诉求推荐策略(没有学习模块)相比，也可以将累积期望遗憾减少 25.71%。

[0046] 实验中，采用 Dropout 进行动作探索的动作选择算法效果要优于不采用 Dropout 的动作选择算法，这是由于采用 Dropout 的动作采样可以近似看成汤普森采样，平衡探索(Exploration)和利用(Exploitation)，更好地在模型的动作空间中进行采样，从而收敛到更好的局部最优解。在 Dropout 比例为 20%，40%，60%，80% 的四组实验中，随着 Dropout 比例的增大，模型的表现先提升后下降。这可能是由于在 Dropout 比例低的时候，模型采用了较保守的探索策略，从而更可能收敛到一个比较差的局部最优解；而在 Dropout 比例较高的时候，模型频繁进行探索，导致其不能充分利用学习到的知识，使得性能下降。其中：当 Dropout 比例为 40% 时，模型的表现现在训练中和收敛后相对于其他 Dropout 比例的模型均取得了较好的效果，这说明通过设置合适的 Dropout 比例来权衡探索(Exploration)和利用

[Exploitation]能够优化模型的表现。

[0047] 如图 3 所示，为交互初期可能会产生累积采纳率下降的情况，这可能是由于训练初期模型的不确定性较大导致的。分析了实验过程中的实时累积期望遗憾和累积采纳率后，本实施例发现在累积采纳率下降的同一时期内，累积期望遗憾的增量明显下降，说明探索

[Exploration]能够使模型更好地学习机器人诉求。

[0048] 为了验证模型的泛化能力，本实施例进行了对照实验。实验中，实验组为 Dropout 比例 40%的模型，对照组为同样的模型，但是本实施例对对照组模型输入中的诉求相关信息进行了随机化处理，本实施例对实验结果进行了与图二中类似的处理，并在图 4 和表 1 中进行了展示。由图 4 和表 1 中的实验结果可以得到，有诉求相关信息输入的模型性能优于没有诉求相关信息输入的模型，这体现了模型可以通过诉求相关信息更好地学习机器人的诉求。

[0049] 现有常规手段并未建立与机器人资源分配满意度的交互过程，并未建模机器人的诉求偏好，并未使用机器人对于资源分配结果的反馈信号来学习机器人的诉求偏好，并未使用在线学习中探索利用的方式向机器人进行个性化诉求推荐的优化，并未实现机器人诉求偏好推荐策略的大规模泛化型应用。

[0050] 与现有技术相比，本方法显著提升了机器人诉求满足率，资源分配效率，推荐策略泛化性。

[0051] 上述具体实施可由本领域技术人员在不背离本发明原理和宗旨的前提下以不同的方式对其进行局部调整，本发明的保护范围以权利要求书为准且不由上述具体实施所限，在其范围内的各个实现方案均受本发明之约束。

# 说明书附图

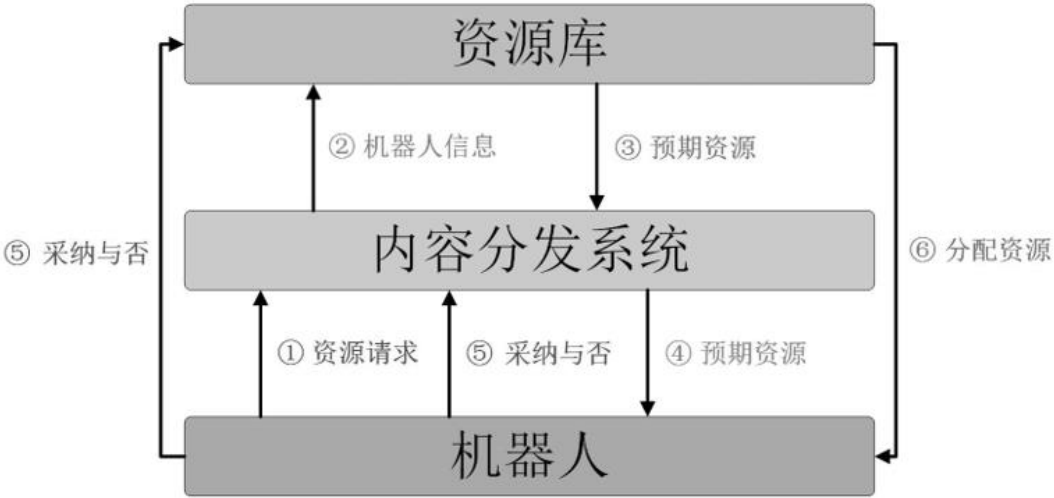


图 1

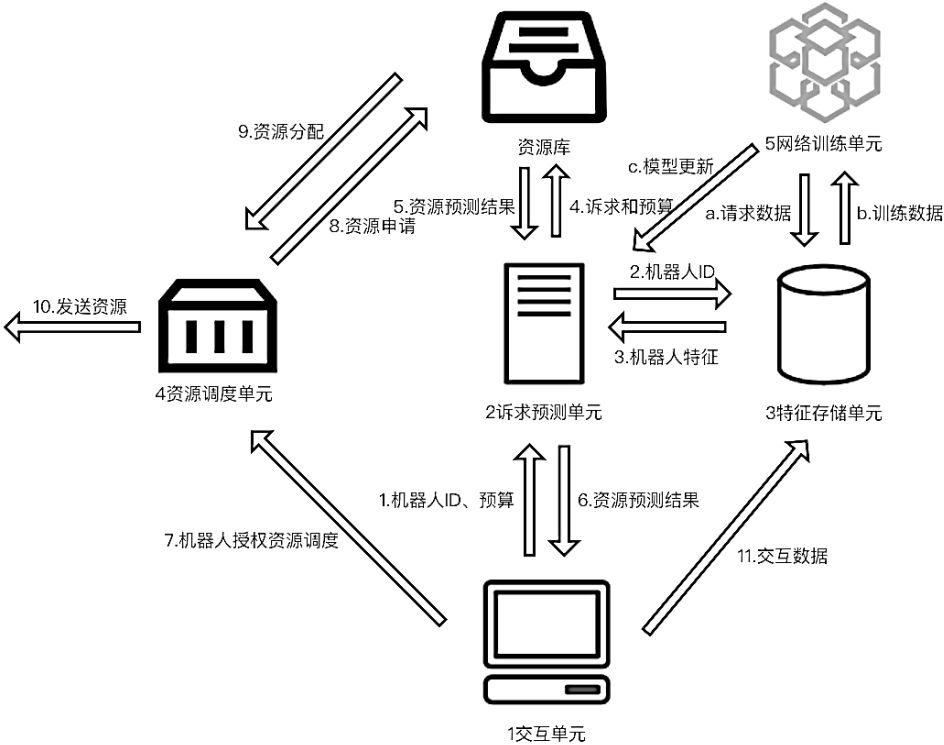


图 2



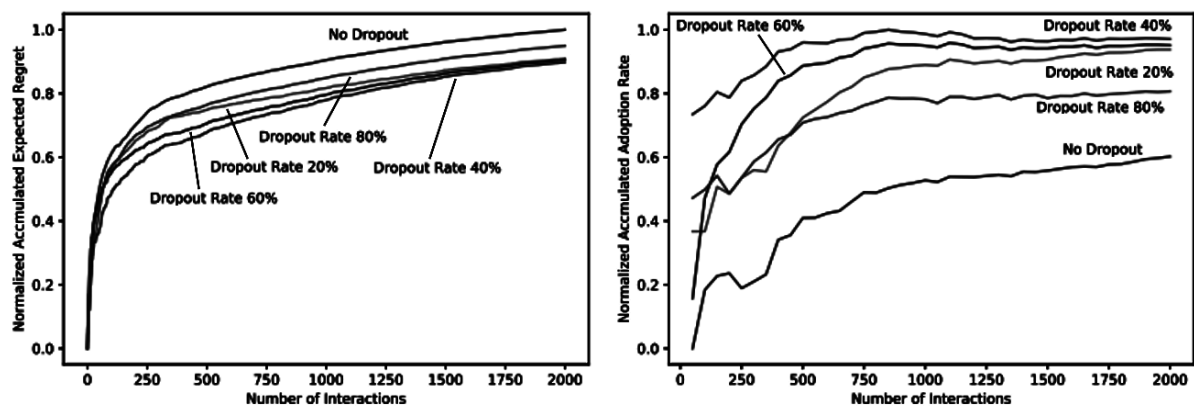


图 2

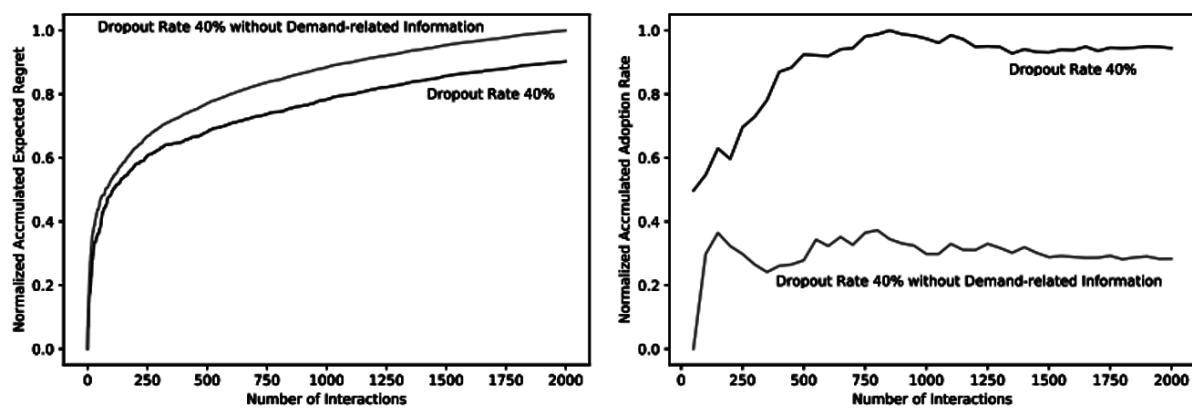


图 3

# 权 利 要 求 书

1、一种工业平台资源优化分配装置，其特征在于，包括：内容分发系统和资源库，其中：内容分发系统生成机器人的资源预测请求并输出至资源库，根据资源库的反馈进行最优资源配置，实现机器人服务流程的同时基于新增的数据更新内容分发系统中诉求预测单元中的神经网络模型，资源库接收内容分发系统发出的资源预测请求，预测潜在可分配的最优资源配置，接收内容分发系统的资源调度单元的资源申请，基于资源申请分配资源。

2、根据权利要求 1 所述的工业平台资源优化分配装置，其特征是，所述的内容分发系统包括：交互单元、诉求预测单元、特征存储单元、资源调度单元和网络训练单元，其中：交互单元接收机器人的资源请求，将机器人 ID 和预算发送给诉求预测单元；诉求预测单元将机器人 ID 发送给特征存储单元；特征存储单元将该机器人特征发送给诉求预测单元；诉求预测单元中的神经网络基于机器人特征预测出机器人的诉求，将诉求与预算发送给资源库；诉求预测单元根据来自资源库的资源预测结果发送给交互单元，交互单元询问机器人是否采纳；并在机器人采纳该资源调度结果时，将机器人授权的资源调度结果发送给资源调度单元；资源调度单元发送资源申请请求给资源库；资源调度单元将资源发送给机器人；该轮交互结束，交互单元将新一轮的交互数据发送给特征存储单元。

3、根据权利要求 2 所述的工业平台资源优化分配装置，其特征是，所述的神经网络模型，通过以下方式进行训练：网络训练单元向特征存储单元发送数据请求；特征存储单元将训练数据发送给网络训练单元；网络训练单元训练神经网络模型，并更新诉求预测单元中的神经网络模型。

4、根据权利要求 1~3 中任一所述装置的工业平台信息优化分发处理方法，其特征在于，基于机器人发起资源申请请求时，内容分发系统从申请请求中解析出机器人相关信息并生成估计的机器人需求，并将机器人需求以及为该机器人分配的预算等信息发送到资源库，询问可分配的资源；资源库根据内容分发系统提供的需求和预算估计能够获取的资源，并将预计的可分配资源  $v = [v_1, v_2, \dots, v_n]^T$  返回，内容分发系统据此向机器人发送资源申请结果，并根据机器人的采纳反馈信号，通过资源库将基于该需求的真实资源结果分配给机器人。

5、根据权利要求 4 所述的工业平台信息优化分发处理方法，其特征是，所述的可分配的

资源是指：在预算等各种约束下，机器人获得的资源结果，具体为： $v = [v_1, v_2, \dots, v_n]^T$ ，其中： $n$ 表示资源的种类数，值 $v_i$ 表示第 $i$ 维资源的量。

6、根据权利要求 4 所述的工业平台信息优化分发处理方法，其特征是，所述的机器人相关信息，包括：机器人拥有的资源申请预算、机器人对于不同的资源的偏好，即诉求权重向量： $w = [w_1, w_2, \dots, w_n]^T$ ，其中： $w_i$ 表示机器人对 $i$ 维报表的偏好权重。

# 说明书摘要

---

一种工业平台资源优化分配装置及方法，包括：内容分发系统和资源库，其中：内容分发系统生成机器人的资源预测请求并输出至资源库，根据资源库的反馈进行最优资源配置，实现机器人服务流程的同时基于新增的数据更新内容分发系统中诉求预测单元中的神经网络模型，资源库接收内容分发系统发出的资源预测请求，预测潜在可分配的最优资源配置，接收内容分发系统的资源调度单元的资源申请，基于资源申请分配资源。本发明通过对机器人诉求和优化目标进行建模，本发明推荐给机器人其可以分配到的资源，并通过机器人的反馈获取资源分配的合理性，打破服务器端与机器人之间信息不对称的僵局。

## 摘要附图

