



(12) 发明专利

(10) 授权公告号 CN 110457577 B

(45) 授权公告日 2023. 04. 14

(21) 申请号 201910618331.4

(51) Int.Cl.

(22) 申请日 2015.04.01

G06F 16/9535 (2019.01)

(65) 同一申请的已公布的文献号

审查员 王雨宁

申请公布号 CN 110457577 A

(43) 申请公布日 2019.11.15

(62) 分案原申请数据

201510151908.7 2015.04.01

(73) 专利权人 创新先进技术有限公司

地址 开曼群岛大开曼岛乔治镇医院路27号

开曼企业中心

(72) 发明人 王丹丹 潘晓峰 徐笛

(74) 专利代理机构 北京国昊天诚知识产权代理

有限公司 11315

专利代理师 朱文杰

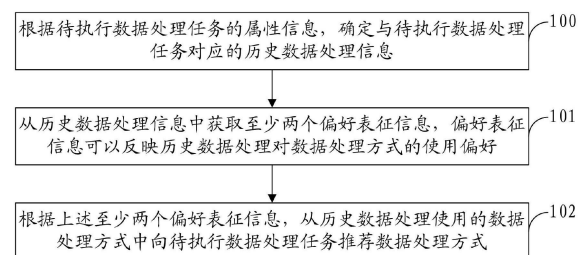
权利要求书3页 说明书12页 附图2页

(54) 发明名称

数据处理方法、装置、设备和计算机存储介质

(57) 摘要

本申请提供一种数据处理方法及装置。方法包括：根据待执行数据处理任务的属性信息，确定与待执行数据处理任务对应的历史数据处理信息；从历史数据处理信息中获取至少两个偏好表征信息，偏好表征信息可以反映历史数据处理对数据处理方式的使用偏好；根据至少两个偏好表征信息，从历史数据处理使用的数据处理方式中向待执行数据处理任务推荐数据处理方式。本申请采用了至少两个可以反映历史数据处理对数据处理方式的使用偏好的信息进行数据处理方式的推荐，可以降低重新选择数据处理方式的概率，有利于提高数据处理的执行效率。



1. 一种数据处理方法,其特征在于,该方法包括:

获取待支付业务对应的已完成支付业务的支付时间、使用的支付方式类型以及支付金额;

根据所述已完成支付业务的支付时间,对所述已完成支付业务使用的支付方式按照类型进行排序,以获得第一排序结果;

根据所述已完成支付业务的支付金额,将所述已完成支付业务使用的支付方式分为第一类和第二类;

对所述第一类中的支付方式按照类型进行排序,以获得第二排序结果,对所述第二类中的数据处理方式按照类型进行排序,以获得第三排序结果;

根据所述第一排序结果、所述第二排序结果和所述第三排序结果,从所述已完成支付业务使用的支付方式中确定向所述待支付业务推荐的支付方式。

2. 根据权利要求1所述的方法,其特征在于,所述获取待支付业务对应的已完成支付业务的支付时间、使用的支付方式类型以及支付金额包括:

根据所述待支付业务的属性信息,确定与所述待支付业务对应的已完成支付业务详情;

从所述已完成支付业务详情中获取所述待支付业务对应的已完成支付业务的支付时间、使用的支付方式类型以及支付金额。

3. 根据权利要求2所述的方法,其特征在于,根据所述待支付业务的属性信息,确定与所述待支付业务对应的已完成支付业务详情包括以下中的一种或任意组合:

获取与所述待支付业务的用户标识相关的已完成支付业务详情;

获取支付金额的值在所述待支付业务的支付金额的值所在区间范围内的已完成支付业务详情;

获取支付时间在所述待支付业务的执行时间所在时间范围内的已完成支付业务详情。

4. 根据权利要求1所述的方法,其特征在于,所述根据所述已完成支付业务的支付时间,对所述已完成支付业务使用的支付方式按照类型进行排序,包括:

根据所述已完成支付业务的支付时间和所述待支付业务的支付时间,获取所述已完成支付业务使用的支付方式对应的推荐权重,其中所述已完成支付业务的支付时间和待支付业务的支付时间之差越大,所述已完成支付业务使用的支付方式对应的推荐权重越小;

根据所述已完成支付业务使用的支付方式对应的推荐权重和所述已完成支付业务使用的支付方式的类型,对所述已完成支付业务使用的支付方式按照类型进行排序;

所述根据所述已完成支付业务的支付金额,将所述已完成支付业务使用的支付方式分为第一类和第二类,包括:

根据所述已完成支付业务的支付金额的分割点,将所述已完成支付业务使用的支付方式分为第一类和第二类;

所述对所述第一类中的支付方式按照类型进行排序,以获得第二排序结果,对所述第二类中的数据处理方式按照类型进行排序,以获得第三排序结果,包括:

根据所述第一类中的支付方式对应的推荐权重和类型,对所述第一类中的支付方式按照类型进行排序,以获得第二排序结果,根据所述第二类的支付方式对应的推荐权重和类型,对所述第二类中的数据处理方式按照类型进行排序,以获得第三排序结果。

5. 根据权利要求3所述的方法,其特征在于,所述已完成支付业务的支付金额的分割点采用如下方式确定:

对所述已完成支付业务的支付金额进行排序;

获取排序后两两相邻的支付金额的均值分别作为候选分割点;

根据所述已完成支付业务使用的支付方式对应的推荐权重和所述已完成支付业务使用的支付方式的类型,计算所述候选分割点的不纯度;

选择不纯度最小的候选分割点作为所述分割点。

6. 根据权利要求5所述的方法,其特征在于,该方法还包括:

获取所述候选分割点的均值与所述分割点之差,作为所述已完成支付业务使用的支付金额的波动系数。

7. 根据权利要求4所述的方法,其特征在于,所述根据所述第一排序结果、所述第二排序结果和所述第三排序结果,从所述已完成支付业务使用的支付方式中确定向所述待支付业务推荐的支付方式,包括:

若所述已完成支付业务的支付金额的波动系数小于预设门限,根据所述第一排序结果,确定向所述待支付业务推荐的支付方式;

否则,根据所述待支付业务的支付金额与所述分割点的大小关系,确定根据所述第二排序结果或所述第三排序结果向所述待支付业务推荐的支付方式。

8. 一种数据处理装置,其特征在于,该装置包括:

获取模块,用于获取待支付业务对应的已完成支付业务的支付时间、使用的支付方式的类型以及支付金额;

推荐模块,用于根据所述已完成支付业务的支付时间,对所述已完成支付业务使用的支付方式按照类型进行排序,以获得第一排序结果;根据所述已完成支付业务的支付金额,将所述已完成支付业务使用的支付方式分为第一类和第二类;对所述第一类中的支付方式按照类型进行排序,以获得第二排序结果,对所述第二类中的数据处理方式按照类型进行排序,以获得第三排序结果;根据所述第一排序结果、所述第二排序结果和所述第三排序结果,从所述已完成支付业务使用的支付方式中确定向所述待支付业务推荐的支付方式。

9. 根据权利要求8所述的装置,其特征在于,该装置还包括:

确定模块,用于根据所述待支付业务的属性信息,确定与所述待支付业务对应的已完成支付业务详情;

所述获取模块,具体用于从所述已完成支付业务详情中获取所述待支付业务对应的已完成支付业务的支付时间、使用的支付方式的类型以及支付金额。

10. 根据权利要求9所述的装置,其特征在于,所述确定模块具体用于执行以下所列操作中的一种或任意结合:

获取与所述待支付业务的用户标识相关的已完成支付业务详情;

获取支付金额的值在所述待支付业务的支付金额的值所在区间范围内的已完成支付业务详情;

获取支付时间在所述待支付业务的执行时间所在时间范围内的已完成支付业务详情。

11. 根据权利要求9所述的装置,其特征在于,所述推荐模块具体用于:

根据所述已完成支付业务的支付时间和所述待支付业务的支付时间,获取所述已完成

支付业务使用的支付方式对应的推荐权重,其中所述已完成支付业务的支付时间和待支付业务的支付时间之差越大,所述已完成支付业务使用的支付方式对应的推荐权重越小;

根据所述已完成支付业务使用的支付方式对应的推荐权重和所述已完成支付业务使用的支付方式的类型,对所述已完成支付业务使用的支付方式按照类型进行排序,以获得第一排序结果;

根据所述已完成支付业务的支付金额的分割点,将所述已完成支付业务使用的支付方式分为第一类和第二类,根据所述第一类中的支付方式对应的推荐权重和类型,对所述第一类中的支付方式按照类型进行排序,以获得第二排序结果,根据所述第二类的支付方式对应的推荐权重和类型,对所述第二类中的数据处理方式按照类型进行排序,以获得第三排序结果。

12. 根据权利要求11所述的装置,其特征在于,所述推荐模块,还用于:

对所述已完成支付业务的支付金额进行排序;

获取排序后两两相邻的支付金额的均值分别作为候选分割点;

根据所述已完成支付业务使用的支付方式对应的推荐权重和所述已完成支付业务使用的支付方式的类型,计算所述候选分割点的不纯度;

选择不纯度最小的候选分割点作为所述分割点。

13. 根据权利要求12所述的装置,其特征在于,所述推荐模块,还用于:

获取所述候选分割点的均值与所述分割点之差,作为所述已完成支付业务使用的支付金额的波动系数。

14. 根据权利要求11所述的装置,其特征在于,所述推荐模块在根据所述第一排序结果、所述第二排序结果和所述第三排序结果,从所述已完成支付业务使用的支付方式中确定向所述待支付业务推荐的支付方式时,具体执行:

若所述已完成支付业务的支付金额的波动系数小于预设门限,根据所述第一排序结果,确定向所述待支付业务推荐的支付方式;

否则,根据所述待支付业务的支付金额与所述分割点的大小关系,确定根据所述第二排序结果或所述第三排序结果向所述待支付业务推荐的支付方式。

15. 一种设备,包括

存储器,包括一个或者多个程序;

一个或者多个处理器,耦合到所述存储器,执行所述一个或者多个程序,以实现如权利要求1至7任一权项所述方法中执行的操作。

16. 一种计算机存储介质,所述计算机存储介质被编码有计算机程序,所述程序在被一个或多个计算机执行时,使得所述一个或多个计算机执行如权利要求1至7任一权项所述方法中执行的操作。

数据处理方法、装置、设备和计算机存储介质

[0001] 本申请是申请日为2015年04月01日,申请号为201510151908.7,发明名称为“数据处理方法及装置”的中国专利申请的分案申请。

【技术领域】

[0002] 本申请涉及互联网技术领域,尤其涉及一种数据处理方法及装置。

【背景技术】

[0003] 随着互联网技术的发展,基于互联网的应用越来越多,并且越来越复杂。通常,应用都会涉及数据处理,一般应用可以使用多个数据处理方式进行数据处理。

[0004] 在现有技术中,一般是根据数据处理方式的历史使用频次,从中选择一个使用频次最高的数据处理方式推荐给待执行数据处理任务。但实际上,使用频次最高的数据处理方式不一定是待执行数据处理任务所需的,这就需要重新选择数据处理方式,影响数据处理的执行效率。

【发明内容】

[0005] 本申请的多个方面提供一种数据处理方法及装置,用以推荐数据处理方式,提高数据处理的执行效率。

[0006] 本申请的一方面,提供一种数据处理方法,包括:

[0007] 根据待执行数据处理任务的属性信息,确定与所述待执行数据处理任务对应的历史数据处理信息;

[0008] 从所述历史数据处理信息中获取至少两个偏好表征信息,所述偏好表征信息可以反映历史数据处理对数据处理方式的使用偏好;

[0009] 根据所述至少两个偏好表征信息,从所述历史数据处理使用的数据处理方式中向待执行数据处理任务推荐数据处理方式。

[0010] 本申请的另一方面,提供一种数据处理装置,包括:

[0011] 确定模块,用于根据待执行数据处理任务的属性信息,确定与所述待执行数据处理任务对应的历史数据处理信息;

[0012] 获取模块,用于从所述历史数据处理信息中获取至少两个偏好表征信息,所述偏好表征信息可以反映历史数据处理对数据处理方式的使用偏好;

[0013] 推荐模块,用于根据所述至少两个偏好表征信息,从所述历史数据处理使用的数据处理方式中向待执行数据处理任务推荐数据处理方式。

[0014] 在本申请中,从历史数据处理信息中,获取至少两个可以反映历史数据处理对数据处理方式的使用偏好的偏好表征信息,根据所获取的至少两个偏好表征信息,从历史数据处理使用的数据处理方式中向待执行数据处理任务推荐数据处理方式,由于同时采用了至少两个可以反映历史数据处理对数据处理方式的使用偏好的信息进行数据处理方式的推荐,所推荐的数据处理方式是符合对数据处理方式的使用偏好的,因此可以向待执行数

据处理任务推荐所需的数据处理方式,这样待执行数据处理任务可以直接基于所推荐的数据处理方式,可以降低重新选择数据处理方式的概率,有利于提高数据处理的执行效率。

【附图说明】

[0015] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍,显而易见地,下面描述中的附图是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0016] 图1为本申请一实施例提供的数据处理方法的流程示意图;

[0017] 图2为本申请一实施例提供的对支付方式进行划分的层次示意图;

[0018] 图3为本申请一实施例提供的数据处理装置的结构示意图。

【具体实施方式】

[0019] 为使本申请实施例的目的、技术方案和优点更加清楚,下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,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,都属于本申请保护的范围。

[0020] 随着应用的发展,应用所能实现的功能越来越强大,也越来越复杂。目前,几乎所有应用都会涉及数据处理,一般一个数据处理过程可以采用多个数据处理方式,但是不同数据处理方式带来的效果截然不同。在数据处理过程中,如果推荐的数据处理方式不适合该数据处理过程,那么就要更换数据处理方式,这会导致数据处理执行效率较低,那么如何推荐更加合适的数据处理方式,以提高数据处理的执行效率呢。针对该问题,本申请提供一种数据处理方法,该方法的主要思想是:基于历史数据处理对数据处理方式的使用偏好,向待执行数据处理任务推荐数据处理方式,这样向待执行数据处理任务所推荐的数据处理方式是符合数据处理对数据处理方式的使用偏好的数据处理方式,符合待执行数据处理任务需求的概率较高,因此对待执行数据处理任务来说,可以提高执行效率。其中,历史数据处理对数据处理方式的使用偏好可以通过历史数据处理在使用数据处理方式时和/或使用过程中的有关信息来体现。为了提高更好的体现历史数据处理在使用数据处理方式时的使用偏好,本申请采用至少两个可以反映历史数据处理在使用数据处理方式时的使用偏好的信息,并且将这种信息称为偏好表征信息。

[0021] 为了使本领域技术人员更好地理解本发明中的技术方案,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详细地描述。

[0022] 图1为本申请一实施例提供的数据处理方法的流程示意图。如图1所示,该方法包括:

[0023] 100、根据待执行数据处理任务的属性信息,确定与待执行数据处理任务对应的历史数据处理信息。

[0024] 101、从历史数据处理信息中获取至少两个偏好表征信息,偏好表征信息可以反映历史数据处理对数据处理方式的使用偏好。

[0025] 102、根据上述至少两个偏好表征信息,从历史数据处理使用的数据处理方式中向待执行数据处理任务推荐数据处理方式。

[0026] 在本申请中,所述历史数据处理是指已经执行并且完成的数据处理过程,可以是近期一段时间内的历史数据,例如可以是近3个月的数据,所述待执行数据处理任务是指即将执行的数据处理过程。所述历史数据处理和所述待执行数据处理任务属于同一类应用场景中的数据处理过程,且对应用场景不做限定。例如,历史数据处理和待执行数据处理任务的应用场景可以是支付业务、下载业务、搜索业务等。另外,本申请的历史数据处理和待执行数据处理任务可以采用多种数据处理方式来实现。根据应用场景类型的不同,需要向待执行数据处理任务推荐的数据处理方式也会有所不同。例如,对于支付类业务,需要推荐的数据处理方式可以是具有支付功能的数据处理方式,例如可以各种信用卡、银行卡、财付通、支付宝等支付方式。

[0027] 具体的,当需要向待执行数据处理任务推荐数据处理方式时,首先根据待执行数据处理任务的属性信息,确定与该待执行数据处理任务对应的历史数据处理信息,接着从历史数据处理信息中获取至少两个可以反映历史数据处理对数据处理方式的使用偏好的偏好表征信息,之后根据所获取的偏好表征信息,从历史数据处理使用的数据处理方式中,向待执行数据处理任务推荐数据处理方式。

[0028] 上述待执行数据处理任务的属性信息包括以下所列中的一种或任意结合:所述待执行数据处理任务的用户标识、所述待执行数据处理任务的数据对象的值、所述待执行数据处理任务的执行时间点。

[0029] 所述根据待执行数据处理任务的属性信息,确定与所述待执行数据处理任务对应的历史数据处理,相应地包括以下所列中的一种或任意结合:

[0030] 获取与所述待执行数据处理任务的用户标识相关的历史数据处理信息;

[0031] 获取数据对象的值在所述待执行数据处理任务的数据对象的值所在区间范围内的历史数据处理信息;

[0032] 获取执行时间在所述待执行数据处理任务的执行时间所在的时间范围内的历史数据处理信息。

[0033] 在本实施例中,从获取的历史数据处理信息中,通过至少两个偏好表征信息来表示历史数据处理在使用数据处理方式时的使用偏好,达到基于历史数据处理在使用数据处理方式时的使用偏好,向待执行数据处理任务推荐数据处理方式的目的,这样向待执行数据处理任务所推荐的数据处理方式是符合数据处理过程对数据处理方式的使用偏好的,符合待执行数据处理任务需求的概率较高,因此对依赖于所推荐的数据处理方式的待执行数据处理任务来说,可以提高该数据处理的执行效率。

[0034] 在一可选实施方式中,历史数据处理信息中可以反映历史数据处理对数据处理方式的使用偏好的偏好表征信息可以包括但不限于:历史数据处理的类型、历史数据处理的执行时间(实际上等价于历史数据处理使用数据处理方式的时间)、历史数据处理使用的数据处理方式的类型以及历史数据处理的数据对象、历史数据处理所处的位置环境等等。

[0035] 本申请发明人经过大量试验发现,在上述偏好表征信息中,历史数据处理的执行时间、所使用的数据处理方式的类型以及历史数据处理的数据对象等信息在反映对数据处理方式的使用偏好上所能发挥的作用更大。

[0036] 基于上述发现,在一种从历史数据处理信息中获取至少两个偏好表征信息的可选实施方式中,可以获取历史数据处理的执行时间、历史数据处理使用的数据处理方式的类型以及历史数据处理的数据对象。

[0037] 在一可选实施方式中,在当前时间之前完成的数据处理过程均属于历史数据处理。考虑到历史数据处理的数量较多,并且执行时间距离当前时间较远的历史数据处理在向待执行数据处理任务推荐数据处理方式上的贡献相对较小,于是可以预先设定一时间范围,从而获取在该时间范围之内执行的数据处理,这样可以降低所使用的历史数据处理的数据量,有利于提高处理速度。所述时间范围可以是距离当前时间3-5天内,或者距离当前时间2周内,或者距离当前时间1-3月内等等。

[0038] 基于上述,上述获取至少两个偏好表征信息具体可以是:获取指定时间范围内的历史数据处理的执行时间、指定时间范围内的历史数据处理使用的数据处理方式的类型以及历史数据处理的数据对象。

[0039] 进一步,在上述获取的历史数据处理的执行时间、历史数据处理使用的数据处理方式的类型以及历史数据处理的数据对象的基础上,向待执行数据处理任务推荐数据处理方式的实施方式包括:

[0040] 根据历史数据处理的执行时间、历史数据处理使用的数据处理方式的类型以及历史数据处理的数据对象,对历史数据处理使用的数据处理方式进行排序;

[0041] 根据排序结果,从历史数据处理使用的数据处理方式中向待执行数据处理任务推荐数据处理方式。

[0042] 一种具体排序方式的处理流程如下:

[0043] 分别给历史数据处理的执行时间、历史数据处理使用的数据处理方式的类型以及历史数据处理的数据对象配置相应的权重系数,根据权重系统对历史数据处理的执行时间、历史数据处理使用的数据处理方式的类型以及历史数据处理的数据对象进行加权平均,从而获得历史数据处理使用的数据处理方式对应的推荐权重;进而根据推荐权重对历史数据处理使用的数据处理方式进行排序;例如可以按照推荐权重由大到小的顺序进行排序,或者按照推荐权重由小到大的顺序进行排序。

[0044] 基于上述实施方式,在获得排序结果后,可以根据排序结果,向待执行数据处理任务推荐数据处理方式。例如,若按照推荐权重由大到小的顺序进行排序,则可以将排在最前面的数据处理方式推荐给待执行数据处理任务。若按照推荐权重由小到大的顺序进行排序,则可以将排在最后面的数据处理方式推荐给待执行数据处理任务。

[0045] 另一种具体排序方式的处理流程如下:

[0046] 首先,根据历史数据处理的执行时间和待执行数据处理任务的执行时间,获取历史数据处理使用的数据处理方式对应的推荐权重,其中,历史数据处理的执行时间和待执行数据处理任务的执行时间之差越大,历史数据处理使用的数据处理方式对应的推荐权重越小;例如,可以根据如下公式(1),计算历史数据处理使用的数据处理方式对应的推荐权重;

$$[0047] \quad f_i = \frac{1}{pre_date - d_i} \quad (1)$$

[0048] 在公式(1)中, f_i 表示第*i*个历史数据处理使用的数据处理方式对应的推荐权重;

pre_date表示待执行数据处理任务的执行时间; d_i 表示第*i*个历史数据处理的执行时间; $i=1,2,\dots,N$, N 是历史数据处理的个数。

[0049] 接着,根据历史数据处理使用的数据处理方式对应的推荐权重和历史数据处理使用的数据处理方式的类型,对历史数据处理使用的数据处理方式按照类型进行排序,以获得第一排序结果。第一排序结果实际上是对历史数据处理使用的数据处理方式的整体排序。

[0050] 例如,可以根据历史数据处理使用的数据处理方式的类型,统计出每种类型下的数据处理方式,再将每种类型下的数据处理方式对应的推荐权重进行累加,获得每种类型对应的推荐权重,根据每种类型对应的推荐权重,对历史数据处理使用的数据处理方式按照类型进行排序。

[0051] 另外,根据历史数据处理的数据对象的分割点,将历史数据处理使用的数据处理方式分为第一类和第二类,根据第一类中的数据处理方式对应的推荐权重和类型,对第一类中的数据处理方式按照类型进行排序,以获得第二排序结果,根据第二类中的数据处理方式对应的推荐权重和类型,对第二类中的数据处理方式按照类型进行排序,以获得第三排序结果。第二排序结果实际上是对数据对象大于分割点的历史数据处理使用的数据处理方式的排序;第三排序结果实际上是对数据对象小于或等于分割点的历史数据处理使用的数据处理方式的排序。

[0052] 具体的,可以将数据对象大于分割点的历史数据处理使用的数据处理方式归为第一类,将数据对象小于或等于分割点的历史数据处理使用的数据处理方式归为第二类。对每一类中的数据处理方式,可以根据数据处理方式的类型,统计出每种类型下的数据处理方式,再将每种类型下的数据处理方式对应的推荐权重进行累加,获得每种类型对应的推荐权重,根据每种类型对应的推荐权重,对该类中的数据处理方式按照类型进行排序。

[0053] 在上述实施方式中,在根据历史数据处理的数据对象的分割点,将历史数据处理使用的数据处理方式分为第一类和第二类之前,可以获取所述分割点。一种获取分割点的方式包括但不限于:

[0054] 对历史数据处理的数据对象进行排序;获取排序后两两相邻的数据对象的均值分别作为候选分割点;根据历史数据处理使用的数据处理方式对应的推荐权重和历史数据处理使用的数据处理方式的类型,计算候选分割点的不纯度;选择不纯度最小的候选分割点作为上述分割点。

[0055] 具体可以采用公式(2),计算候选分割点。

$$[0056] \quad mean_j = \frac{amt_j \pm amt_{j+1}}{2} \quad (2)$$

[0057] 在公式(2)中, $mean_j$ 表示第*j*个候选分割点; amt_j 表示第*j*个数据对象, amt_{j+1} 表示第*j+1*个数据对象, $j=1,2,\dots,(N-1)$, N 是历史数据处理的个数,也就是数据对象的个数。根据两个相邻数据对象的大小选择“+”或“-”。

[0058] 其中,计算候选分割点的不纯度的方式包括:

[0059] 对每个候选分割点,将历史数据处理的数据对象分别与该候选分割点进行比较,从而将历史数据处理的数据对象分成两类,一类是大于候选分割点的数据对象,记为A类,一类是小于或等于候选分割点的数据对象,记为B类。

[0060] 对于A类,统计该类中数据对象对应的推荐权重之和,作为该类对应的推荐权重,再计算该类对应的推荐权重占总推荐权重的比值,作为时间影响系数。由于推荐权重仅涉及待执行数据处理任务与历史数据处理之间的时间之差,所以时间影响系数反映了时间因素在推荐过程中的作用。另外,A类中的数据对象可能属于一个或多个类型,统计每个类型下的数据对象对应的推荐权重之和,作为该类下每个类型对应的推荐权重,对每个类型,可以根据如下公式(3),计算该类型对应的比值,再将该类下所有类型对应的比值进行累加,作为A类的类型影响系数;由于是从类型的角度对推荐权重进行处理,因此类型影响系数反映了数据处理方式类型这一因素在推荐过程中的作用。

[0061]
$$gini_l = (\text{type_sum}_l / \text{split_sum}_A) \times (1 - \text{type_sum}_l / \text{split_sum}_A) \quad (3)$$

[0062] 在公式(3), $gini_l$ 表示类型 l 对应的比值;表示类型 l 下所有数据对象对应的推荐权重之和,即 $\text{type_sum}_l = \sum f_l$; split_sum_A 表示A类中数据对象对应的推荐权重之和。其中,A类的类型影响系数可表示为 $\sum_A gini_l$ 。 $l = 1, 2, \dots, M$, M 是数据处理方式类型的最大个数。

[0063] 对于B类,统计该类中数据对象对应的推荐权重之和,作为该类对应的推荐权重,再计算该类对应的推荐权重占总推荐权重的比值,作为时间影响系数。由于推荐权重仅涉及待执行数据处理任务与历史数据处理之间的时间之差,所以时间影响系数反映了时间因素在推荐过程中的作用。另外,B类中的数据对象可能属于一个或多个类型,统计每个类型下的数据对象对应的推荐权重之和,作为该类下每个类型对应的推荐权重,对每个类型,可以根据上述公式(3),计算该类型对应的比值,再将该类下所有类型对应的比值进行累加,作为B类的类型影响系数;由于是从类型的角度对推荐权重进行处理,因此类型影响系数反映了数据处理方式类型这一因素在推荐过程中的作用。

[0064] 之后,获取A类的时间影响系数和类型影响系数的乘积,获取B类的时间影响系数和类型影响系数的乘积,再将两个乘积相加作为候选分割点的不纯度。

[0065] 当采用上述方法计算出所有候选分割点的不纯度之后,从中选择最小不纯度对应的候选分割点作为上述使用的分割点。

[0066] 在一可选实施方式中,还可以获取各候选分割点的均值,将候选分割点的均值与上述分割点之差,作为历史数据处理的数据对象的波动系数。例如,该波动系数可以是基尼系数。

[0067] 基于上述排序结果,从历史数据处理使用的数据处理方式中向待执行数据处理任务推荐数据处理方式的方式包括:

[0068] 将历史数据处理的数据对象的波动系数与预设门限进行比较;

[0069] 若上述历史数据处理的数据对象的波动系数小于预设门限,说明数据对象的大小相对较平稳,对推荐过程的影响相对较小,则在向待执行数据处理任务推荐数据处理方式时,可以不考虑数据对象这一因素的影响,则根据第一排序结果,向待执行数据处理任务推荐数据处理方式;

[0070] 若上述历史数据处理的数据对象的波动系数大于或等于预设门限,说明数据对象大小的变化较大,对推荐过程的影响相对较大,则在向待执行数据处理任务推荐数据处理方式时,应该考虑数据对象这一因素的影响,于是根据待执行数据处理任务的数据对象与分割点的关系,确定根据第二排序结果或第三排序结果向待执行数据处理任务推荐数据处

理方式。

[0071] 当待执行数据处理任务的大于分割点时,确定采用第二排序结果向待执行数据处理任务推荐数据处理方式;当待执行数据处理任务的数据对象小于或等于分割点时,确定采用第三排序结果向待执行数据处理任务推荐数据处理方式。

[0072] 由上述可见,在向待执行数据处理任务推荐数据处理方式时,除了考虑历史数据处理的执行时间、所使用的数据处理方式的类型之外,还可以根据历史数据处理的数据对象进行区分,如果数据对象影响较大,则考虑数据对象的影响,反之,不考虑数据对象的影响,能够进一步提高所推荐数据处理方式的准确性。

[0073] 进一步,考虑到应用场景较多,为了便于本申请技术方案能够更加充分的发挥其优势,可以预先确定本申请技术方案适用的应用场景。基于此,在根据待执行数据处理任务的属性信息,确定与待执行数据处理任务对应的历史数据处理信息之前,可以判断待执行数据处理任务的应用场景是否属于预设的应用场景;若判断结果为是,则执行根据待执行数据处理任务的属性信息,确定与待执行数据处理任务对应的历史数据处理信息及后续操作。若判断结果为否,则可以不采用本申请技术方案,例如可以采用现有方案向待执行数据处理任务推荐数据处理方式。

[0074] 为便于本领域技术人员更加清楚的理解本申请技术方案,下面将结合具体应用场景进行说明。在下面实施例,将以历史数据处理和待执行数据处理任务均为支付业务中的支付过程为例,上述数据处理方式具体为完成支付流程需要使用的支付方式,例如可以是信用卡、银行卡、支付宝等支付方式;相应的,支付业务涉及的支付金额可以作为上述数据对象。

[0075] 随着互联网的飞速发展,网上支付已成为越来越多网民在支付时的首选。目前,越来越多的支付应用开始大力推广极简收银台,在极简收银台上只显示一种支付方式。当支付方式不是用户想使用的类型时,需要去做切换,这会影响支付效率,用户体验也不好。

[0076] 对于上述问题,本实施例综合考虑已完成支付业务的支付时间、支付操作使用的支付方式的类型以及支付金额等多种可以反映用户在进行支付业务时对支付方式的使用偏好的因素,从而向待支付业务推荐合理或所需的支付方式,从而提高支付效率。

[0077] 具体的,可以采集用户在一定时间内(例如最近3个月)内通过极简收银台已完成的支付业务的详情,包括每笔支付业务的支付时间、每笔支付业务涉及的用户ID、每笔支付业务使用的支付方式、以及每笔支付业务的支付金额等。

[0078] 可选的,对在近3个月内已完成的支付业务所使用的支付方式的类型进行计数;若计数结果等于1,说明近3个月内已完成的支付业务均使用相同类型的支付方式,例如均使用同一张银行卡进行支付,则可以直接将该支付方式推荐给待支付业务;若计数结果大于1,则可以根据近3个月内已完成的支付业务的支付时间、使用的支付方式的类型、涉及的支付金额等信息,向待支付业务推荐支付方式。

[0079] 下面结合图2所示结构示意图,详细说明在上述计数结果大于1时,向待支付业务推荐支付方式的实施过程:

[0080] 假设,近3个月内已完成支付业务一共有N笔,每笔已完成支付业务的支付时间记为d,例如第i笔已完成支付业务的支付时间为 d_i ,每笔已完成支付业务的支付金额记为amt,例如第i笔已完成支付业务的支付金额为 amt_i ,其中, $i=1,2,\dots,N$;将待支付业务的

支付时间记为pre_date。

[0081] 第一步,根据上述公式(1),计算出每笔支付业务使用的支付方式的推荐权重。

[0082] 第二步,计算支付金额的分割点,该步骤具体包括以下操作:

[0083] 1) 将近3个月内已完成支付业务的支付金额进行排序,再根据上述公式(2),计算出N-1个候选分割点,将候选分割点记为mean,例如第j个候选分割点为mean_j。

[0084] 2) 将N笔已完成支付业务分别以每个候选分割点为基准进行归类。

[0085] 具体的,将支付金额大于候选分割点的已完成业务划分为一类,记为upper;将支付金额小于或等于候选分割点的已完成业务划分为另一类,记为lower。例如,第j个候选分割点划分为的两类分别为upper_j和lower_j。

[0086] 另外,计算出所有已完成支付业务使用的支付方式对应的推荐权重的加和,作为总推荐权重,记为mean_sum。总推荐权重对每个候选分割点来说均相同。

[0087] 3) 对每个候选分割点下的每个分类,获取时间影响系数。

[0088] 以第j个候选分割点划分出的upper_j为例,假设该upper_j有m笔已完成支付业务,则根据公式(4),计算出upper_j对应的推荐权重之和;再根据公式(5),计算时间影响系数。

[0089] $split_sum_j = \sum_m f_k$ (4)

[0090] $sigtime_j = \frac{split_sum_j}{mean_sum}$ (5)

[0091] 同理,计算出lower_j的时间影响系数。

[0092] 进一步,计算出其他候选分割点所分割出的两个类的时间影响系数。

[0093] 4) 对每个候选分割点下的每个分类,获取类型影响系数。

[0094] 以第j个候选分割点划分出的upper_j为例,假设该upper_j有m笔已完成支付业务,包括L个支付类型,则根据公式(6),计算出每个支付类型对应的推荐权重之和,再根据公式(3),计算出每个支付类型对应的比值;进一步,根据公式(7),计算出upper_j的类型影响系数。

[0095] $type_sum_l = \sum_x f_l$ (6)

[0096] $sigma_j = sum(gini_l)$ (7)

[0097] 在公式(6)中,假设类型l下的支付方式的个数为x,公式(6)具体为对类型l下的x个支付方式对应的推荐权重求和。对于不同的类型,其下的支付方式的个数会有所不同。

[0098] 同理,计算出lower_j的类型影响系数。

[0099] 进一步,计算出其他候选分割点所分割出的两个类的类型影响系数。

[0100] 5) 计算每个候选分割点的不纯度。

[0101] 以第j个候选分割点为例,则可以根据公式(8),计算其不纯度。

[0102] $mean_impurity_j = \sum sigtime_j \times sigma_j$ (8)

[0103] 第三步,从所有候选分割点中,选择不纯度最小的候选分割点作为金额分割点。

[0104] 第四步,获取波动系数,并对已完成支付业务使用的支付方式进行排序,并输出排序结果、波动系数以及金额分割点等信息。

[0105] 对已完成支付业务使用的支付方式进行排序包括:根据已完成支付业务使用的支付方式对应的推荐权重和支付方式的类型,对已完成支付业务使用的支付方式按照类型进行排序,以获得第一排序结果(即整体排序结果);以及根据金额分割点,将已完成支付业务

使用的支付方式分为第一类和第二类(支付金额高于金额分割点的已完成支付业务使用的支付方式归为第一类,支付金额小于金额分割点的已完成支付业务使用的支付方式归为第二类),根据第一类中的支付方式对应的推荐权重和类型,对第一类中的支付方式按照类型进行排序,以获得第二排序结果(高于金额分割点的排序结果),根据第二类中的支付方式对应的推荐权重和类型,对第二类中的支付方式按照类型进行排序,以获得第三排序结果(低于金额分割点的排序结果)。

[0106] 进一步,上述输出排序结果、波动系数以及金额分割点等信息的方式可以是:以用户ID作为主键,将上述金额分割点、第一排序结果、第二排序结果、第三排序结果以及波动系数等作为一条数据记录输出。

[0107] 之后,基于上述输出结果,向待支付业务推荐支付方式。

[0108] 具体推荐过程如下:

[0109] 将上述波动系数与预设的门限进行比较;所述预设的门限可以是但不限于:0.5;

[0110] 若上述波动系数小于预设的门限,说明支付金额相对较平稳,对推荐过程的影响相对较小,则在向待支付业务推荐支付方式时,可以不考虑支付金额这一因素的影响,则根据第一排序结果(即整体排序结果),向待支付业务推荐支付方式;

[0111] 若上述波动系数大于或等于预设的门限,说明支付金额变化比较大,对推荐过程的影响相对较大,则在向待支付业务推荐支付方式时,需要考虑支付金额这一因素的影响,则根据待支付业务的支付金额与金额分割点的大小关系,确定是根据第二排序结果(即高于金额分割点的排序结果)还是根据第三排序结果(即低于金额分割点的排序结果),向待支付业务推荐支付方式。具体的,当待支付业务的支付金额大于金额分割点时,采用第二排序结果向待支付业务推荐支付方式;当待支付业务的支付金额小于金额分割点时,采用第三排序结果向待支付业务推荐支付方式。

[0112] 发明人经过大量实验得出:采用本实施例提供的方法通过极简收银台向待支付业务推荐支付方式的准确率高达87%,而采用现有技术默认向待支付业务推荐上一次使用的支付方式的准确率为79%,本实施例比之前方法在准确率上提升8个百分点。

[0113] 由上述可见,本申请提供的技术方案应用于支付场景中,通过采集用户在极简收银台上指定时间内已完成的支付业务的支付时间、使用的支付方式的类型、涉及到的支付金额等多维度的信息,通过这些信息来反映用户在极简收银台上进行支付业务时对支付方式的使用习惯,最终输出向待支付业务进行推荐所需的信息,基于所输出的信息向待支付业务推荐支付方式,这种推荐方式可以充分尊重用户在使用支付方式时的使用习惯,减少支付方式的切换,从而提高支付业务的效率,优化用户体验。

[0114] 需要说明的是,对于前述的各方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本申请并不受所描述的动作顺序的限制,因为依据本申请,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作和模块并不一定是本申请所必须的。

[0115] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中未详述的部分,可以参见其他实施例的相关描述。

[0116] 图3为本申请一实施例提供的数据处理装置的结构示意图。如图3所示,该装置包

括：确定模块30、获取模块31和推荐模块32。

[0117] 确定模块30，用于根据待执行数据处理任务的属性信息，确定与所述待执行数据处理任务对应的历史数据处理信息。

[0118] 获取模块31，用于从确定模块30所确定的历史数据处理信息中获取至少两个偏好表征信息，偏好表征信息可以反映历史数据处理对数据处理方式的使用偏好。

[0119] 推荐模块32，用于根据获取模块31获取的至少两个偏好表征信息，从历史数据处理使用的数据处理方式中向待执行数据处理任务推荐数据处理方式。

[0120] 可选的，上述待执行数据处理任务的属性信息包括以下所列中的一种或任意结合：待执行数据处理任务的用户标识、待执行数据处理任务的数据对象的值、待执行数据处理任务的执行时间；

[0121] 相应的，确定模块30具体用于执行以下所列操作中的一种或任意结合：

[0122] 获取与待执行数据处理任务的用户标识相关的历史数据处理信息；

[0123] 获取数据对象的值在待执行数据处理任务的数据对象的值所在区间范围内的历史数据处理信息；

[0124] 获取执行时间在所述待执行数据处理任务的执行时间所在的时间范围内的历史数据处理信息。

[0125] 在一可选实施方式中，获取模块31具体用于：从历史数据处理信息中，获取历史数据处理的执行时间、历史数据处理使用的数据处理方式的类型以及历史数据处理的数据对象。

[0126] 在一可选实施方式中，推荐模块32具体用于：根据历史数据处理的执行时间、历史数据处理使用的数据处理方式的类型以及历史数据处理的数据对象，对历史数据处理使用的数据处理方式进行排序；

[0127] 根据排序结果，从历史数据处理使用的数据处理方式中向待执行数据处理任务推荐数据处理方式。

[0128] 进一步，推荐模块32在对历史数据处理使用的数据处理方式进行排序时，具体可用于：

[0129] 根据历史数据处理的执行时间和待执行数据处理任务的执行时间，获取历史数据处理使用的数据处理方式对应的推荐权重，其中，历史数据处理的执行时间和待执行数据处理任务的执行时间之差越大，历史数据处理使用的数据处理方式对应的推荐权重越小；

[0130] 根据历史数据处理使用的数据处理方式对应的推荐权重和历史数据处理使用的数据处理方式的类型，对历史数据处理使用的数据处理方式按照类型进行排序，以获得第一排序结果；

[0131] 根据历史数据处理的数据对象的分割点，将历史数据处理使用的数据处理方式分为第一类和第二类，根据第一类中的数据处理方式对应的推荐权重和类型，对第一类中的数据处理方式按照类型进行排序，以获得第二排序结果，根据第二类中的数据处理方式对应的推荐权重和类型，对第二类中的数据处理方式按照类型进行排序，以获得第三排序结果。

[0132] 在一可选实施方式中，推荐模块32在对历史数据处理使用的数据处理方式进行排序过程中，还用于：

[0133] 根据历史数据处理的数据对象的大小,对历史数据处理的数据对象进行排序;

[0134] 获取排序后两两相邻的数据对象的均值分别作为候选分割点;

[0135] 根据历史数据处理使用的数据处理方式对应的推荐权重和历史数据处理使用的数据处理方式的类型,计算候选分割点的不纯度;

[0136] 选择不纯度最小的候选分割点作为分割点。

[0137] 在一可选实施方式中,推荐模块32在对历史数据处理使用的数据处理方式进行排序过程中,还用于:获取候选分割点的均值与分割点之差,作为历史数据处理的数据对象的波动系数。

[0138] 在一可选实施方式中,推荐模块32在向待执行数据处理任务推荐数据处理方式时,具体用于:

[0139] 若历史数据处理的数据对象的波动系数小于预设门限,根据第一排序结果,向待执行数据处理任务推荐数据处理方式;

[0140] 若历史数据处理的数据对象的波动系数大于或等于预设门限,根据待执行数据处理任务的数据对象与分割点的关系,确定根据第二排序结果或第三排序结果向待执行数据处理任务推荐数据处理方式。

[0141] 在一可选实施方式中,确定模块30具体用于:判断待执行数据处理任务的应用场景是否属于预设的应用场景,在判断结果为是时,执行根据待执行数据处理任务的属性信息,确定与待执行数据处理任务对应的历史数据处理信息的操作。所述预设的应用场景是指采用本实施例提供的装置进行数据处理能够产生一定有益效果的应用场景,或者简单认为是适于采用本实施例提供的装置进行数据处理的应用场景。

[0142] 本实施例提供的数据处理装置,从历史数据处理信息中,获取至少两个可以反映历史数据处理对数据处理方式的使用偏好的偏好表征信息,根据所获取的至少两个偏好表征信息,从历史数据处理使用的数据处理方式中向待执行数据处理任务推荐数据处理方式,由于同时采用了至少两个可以反映历史数据处理对数据处理方式的使用偏好的信息进行数据处理方式的推荐,所推荐的数据处理方式是符合对数据处理方式的使用偏好的,因此可以向待执行数据处理任务推荐所需的数据处理方式,这样待执行数据处理任务可以直接基于所推荐的数据处理方式进行,可以降低重新选择数据处理方式的概率,有利于提高数据处理的执行效率。

[0143] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统,装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0144] 在本申请所提供的几个实施例中,应该理解到,所揭露的系统,装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0145] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目

的。

[0146] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能单元的形式实现。

[0147] 上述以软件功能单元的形式实现的集成的单元,可以存储在一个计算机可读取存储介质中。上述软件功能单元存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,数据处理方式器,或者网络设备等)或处理器(processor)执行本申请各个实施例所述方法的部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(Read-Only Memory,ROM)、随机存取存储器(Random Access Memory,RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0148] 最后应说明的是:以上实施例仅用以说明本申请的技术方案,而非对其限制;尽管参照前述实施例对本申请进行了详细的说明,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;而这些修改或者替换,并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围。

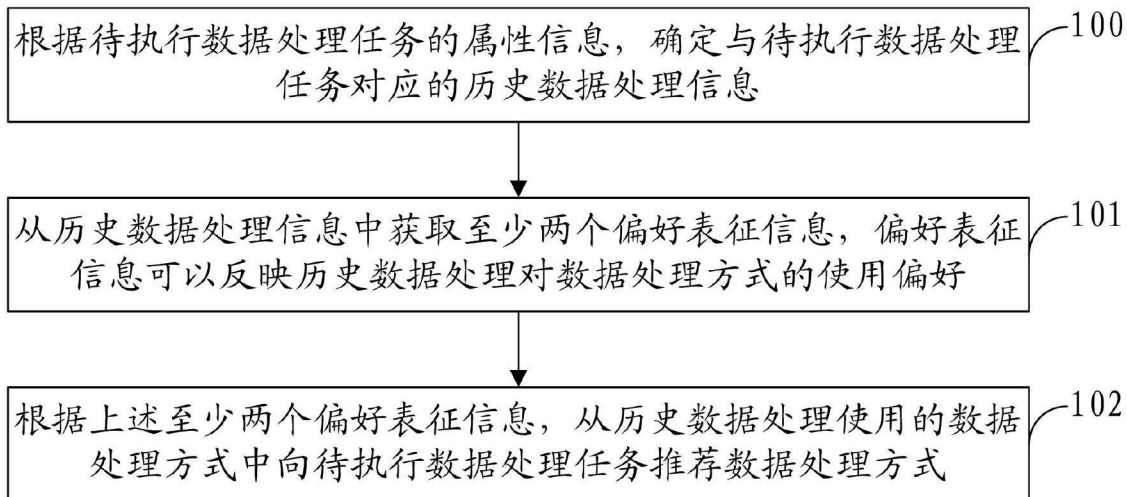


图1

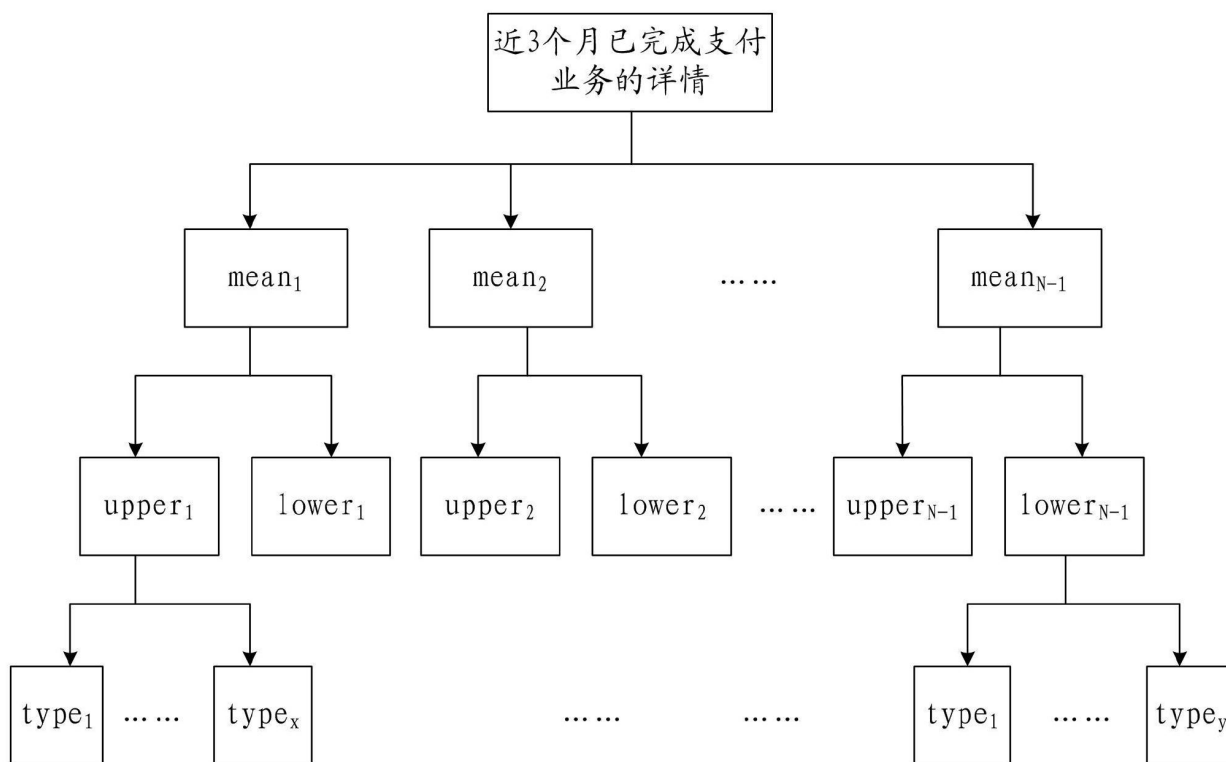


图2

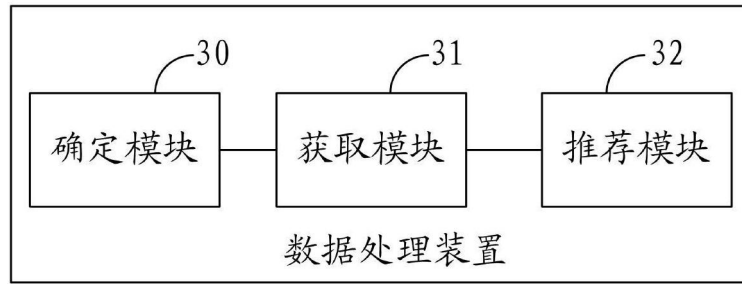


图3