



(12)发明专利申请

(10)申请公布号 CN 106372674 A

(43)申请公布日 2017. 02. 01

(21)申请号 201610873881.7

(22)申请日 2016.09.30

(71)申请人 百度在线网络技术(北京)有限公司

地址 100085 北京市海淀区上地十街10号

百度大厦

(72)发明人 王超

(74)专利代理机构 北京鸿德海业知识产权代理

事务所(普通合伙) 11412

代理人 袁媛

(51)Int.Cl.

G06K 9/62(2006.01)

G06Q 10/02(2012.01)

G06Q 50/30(2012.01)

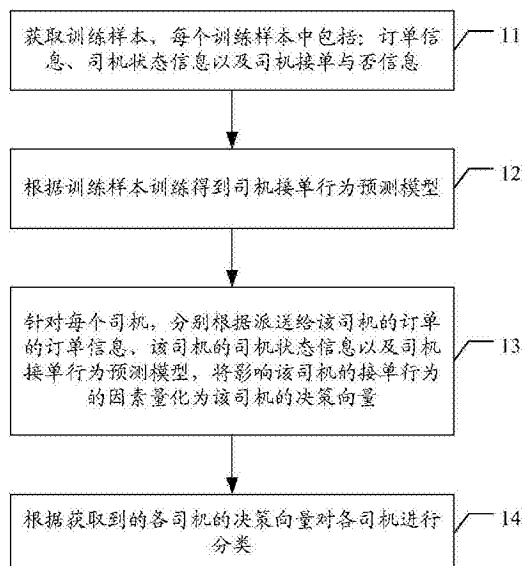
权利要求书4页 说明书9页 附图3页

(54)发明名称

在线叫车服务平台中的司机分类方法和装置

(57)摘要

本发明公开了在线叫车服务平台中的司机分类方法和装置,所述方法包括:获取训练样本,每个训练样本中包括:订单信息、司机状态信息以及司机接单与否信息;根据训练样本训练得到司机接单行为预测模型;针对每个司机,分别根据派送给该司机的订单的订单信息、该司机的司机状态信息以及司机接单行为预测模型,将影响该司机的接单行为的因素量化为该司机的决策向量;根据获取到的各司机的决策向量对各司机进行分类。应用本发明所述方案,能够提高分类结果的准确性等。



1. 一种在线叫车服务平台中的司机分类方法,其特征在于,包括:

获取训练样本,每个训练样本中包括:订单信息、司机状态信息以及司机接单与否信息,并根据所述训练样本训练得到司机接单行为预测模型;

针对每个司机,分别根据派送给所述司机的订单的订单信息、所述司机的司机状态信息以及所述司机接单行为预测模型,将影响所述司机的接单行为的因素量化为所述司机的决策向量;

根据获取到的各司机的决策向量对各司机进行分类。

2. 根据权利要求1所述的方法,其特征在于,

所述订单信息为过往派送的任一订单的订单信息,所述司机状态信息为被派送所述订单的任一司机在所述订单发生时的司机状态信息;

所述司机接单行为预测模型包括:决策树模型。

3. 根据权利要求2所述的方法,其特征在于,

所述将影响所述司机的接单行为的因素量化为所述司机的决策向量包括:

针对所述决策树模型中的每棵决策树,分别进行以下处理:

对于最近预定时长内派送给所述司机的每个订单,分别根据所述订单的订单信息和所述司机的司机状态信息,确定出所述订单的因素向量;

根据所述最近预定时长内的各订单的因素向量以及各订单的接单与否信息,确定出所述决策树对应的决策向量;

根据所述决策树模型中的各决策树对应的决策向量确定出所述司机的决策向量。

4. 根据权利要求3所述的方法,其特征在于,

所述确定出所述订单的因素向量包括:

确定出根据所述订单信息和所述司机状态信息,利用所述决策树模型对所述司机的接单行为进行决策的过程中,在所述决策树上走过的路径;

分别获取所述路径上的每个非叶子节点的属性收益向量 $\text{vec}(t)$, $\text{vec}(t) = (\text{Gain}(1, t), \text{Gain}(2, t), \dots, \text{Gain}(i, t), \dots, \text{Gain}(N, t))$;

其中, t 表示所述路径中的任一非叶子节点,所述 N 表示属性个数,所述属性个数为所述订单信息中包括的信息数与所述司机状态信息中包括的信息数之和,所述订单信息中包括的每条信息以及所述司机状态信息中包括的每条信息分别为一个属性, $\text{Gain}(i, t)$ 表示非叶子节点 t 上分裂时第 i 个属性的纯净度收益;

将所述路径上的各非叶子节点的属性收益向量相加,用相加之和除以所述路径上的非叶子节点数,得到所述订单的因素向量。

5. 根据权利要求3所述的方法,其特征在于,

所述根据所述最近预定时长内的各订单的因素向量以及各订单的司机接单与否信息,确定出所述决策树对应的决策向量包括:

分别计算所述最近预定时长内所述司机接受的每个订单的因素向量与对应的加权系数的乘积,并将各乘积相加,用相加之和除以所述最近预定时长内所述司机接受的订单数,得到接单因素向量;

分别计算所述最近预定时长内所述司机拒接的每个订单的因素向量与对应的加权系数的乘积,并将各乘积相加,用相加之和除以所述最近预定时长内所述司机拒接的订单数,

得到拒单因素向量；

用所述接单因素向量和所述拒单因素向量组成所述决策树对应的决策向量。

6. 根据权利要求5所述的方法，其特征在于，

该方法进一步包括：

在计算所述接单因素向量时，将每个加权系数的取值均设置为1，或者，针对所述司机接受的每个订单，分别将根据所述决策树模型预测出来的所述司机的接单概率作为所述订单的因素向量对应的加权系数；

在计算所述拒单因素向量时，将每个加权系数的取值均设置为1，或者，针对所述司机拒接的每个订单，分别将根据所述决策树模型预测出来的所述司机的拒单概率作为所述订单的因素向量对应的加权系数。

7. 根据权利要求5所述的方法，其特征在于，

所述根据所述决策树模型中的各决策树对应的决策向量确定出所述司机的决策向量包括：

将所述决策树模型中的各决策树对应的决策向量相加，并用相加之和除以所述决策树模型中的决策树数，得到所述司机的决策向量。

8. 根据权利要求7所述的方法，其特征在于，

所述根据获取到的各司机的决策向量对各司机进行分类包括：

对各司机的决策向量进行聚类，将聚类得到的每一簇中的决策向量对应的司机作为一个司机分类；

或者，对各司机的决策向量中的接单因素向量进行聚类，将聚类得到的每一簇中的接单因素向量对应的司机作为一个司机分类；

或者，对各司机的决策向量中的拒单因素向量进行聚类，将聚类得到的每一簇中的拒单因素向量对应的司机作为一个司机分类。

9. 一种在线叫车服务平台中的司机分类装置，其特征在于，包括：模型训练单元、确定单元以及分类单元；

所述模型训练单元，用于获取训练样本，每个训练样本中包括：订单信息、司机状态信息以及司机接单与否信息，并根据所述训练样本训练得到司机接单行为预测模型，将所述司机接单行为预测模型发送给所述确定单元；

所述确定单元，用于针对每个司机，分别根据派送给所述司机的订单的订单信息、所述司机的司机状态信息以及所述司机接单行为预测模型，将影响所述司机的接单行为的因素量化为所述司机的决策向量，并将所述决策向量发送给所述分类单元；

所述分类单元，用于根据获取到的各司机的决策向量对各司机进行分类。

10. 根据权利要求9所述的装置，其特征在于，

所述订单信息为过往派送的任一订单的订单信息，所述司机状态信息为被派送所述订单的任一司机在所述订单发生时的司机状态信息；

所述司机接单行为预测模型包括：决策树模型。

11. 根据权利要求10所述的装置，其特征在于，

所述确定单元中包括：第一处理子单元和第二处理子单元；

所述第一处理子单元，用于针对每个司机，分别按照以下方式获取所述决策树模型中

的每棵决策树对应的决策向量:对于最近预定时长内派送给所述司机的每个订单,分别根据所述订单的订单信息和所述司机的司机状态信息,确定出所述订单的因素向量,根据所述最近预定时长内的各订单的因素向量以及各订单的接单与否信息,确定出所述决策树对应的决策向量;分别将各决策树对应的决策向量发送给所述第二处理子单元;

所述第二处理子单元,用于针对每个司机,分别根据获取到的所述司机对应的、所述决策树模型中的各决策树对应的决策向量,确定出所述司机的决策向量,并发送给所述分类单元。

12. 根据权利要求11所述的装置,其特征在于,

所述第一处理子单元确定出根据所述订单信息和所述司机状态信息,利用所述决策树模型对所述司机的接单行为进行决策的过程中,在所述决策树上走过的路径;分别获取所述路径上的每个非叶子节点的属性收益向量 $\text{vec}(t)$, $\text{vec}(t) = (\text{Gain}(1,t), \text{Gain}(2,t), \dots, \text{Gain}(i,t), \dots, \text{Gain}(N,t))$, 其中, t 表示所述路径中的任一非叶子节点,所述 N 表示属性个数,所述属性个数为所述订单信息中包括的信息数与所述司机状态信息中包括的信息数之和,所述订单信息中包括的每条信息以及所述司机状态信息中包括的每条信息分别为一个属性, $\text{Gain}(i,t)$ 表示非叶子节点 t 上分裂时第 i 个属性的纯净度收益;将所述路径上的各非叶子节点的属性收益向量相加,用相加之和除以所述路径上的非叶子节点数,得到所述订单的因素向量。

13. 根据权利要求11所述的装置,其特征在于,

所述第一处理子单元分别计算所述最近预定时长内所述司机接受的每个订单的因素向量与对应的加权系数的乘积,并将各乘积相加,用相加之和除以所述最近预定时长内所述司机接受的订单数,得到接单因素向量;分别计算所述最近预定时长内所述司机拒接的每个订单的因素向量与对应的加权系数的乘积,并将各乘积相加,用相加之和除以所述最近预定时长内所述司机拒接的订单数,得到拒单因素向量;用所述接单因素向量和所述拒单因素向量组成所述决策树对应的决策向量。

14. 根据权利要求13所述的装置,其特征在于,

所述第一处理子单元进一步用于,

在计算所述接单因素向量时,将每个加权系数的取值均设置为1,或者,针对所述司机接受的每个订单,分别将根据所述决策树模型预测出来的所述司机的接单概率作为所述订单的因素向量对应的加权系数;

在计算所述拒单因素向量时,将每个加权系数的取值均设置为1,或者,针对所述司机拒接的每个订单,分别将根据所述决策树模型预测出来的所述司机的拒单概率作为所述订单的因素向量对应的加权系数。

15. 根据权利要求13所述的装置,其特征在于,

所述第二处理子单元将获取到的所述司机对应的、所述决策树模型中的各决策树对应的决策向量相加,并用相加之和除以所述决策树模型中的决策树数,得到所述司机的决策向量。

16. 根据权利要求15所述的装置,其特征在于,

所述分类单元对各司机的决策向量进行聚类,将聚类得到的每一簇中的决策向量对应的司机作为一个司机分类;

或者,所述分类单元对各司机的决策向量中的接单因素向量进行聚类,将聚类得到的每一簇中的接单因素向量对应的司机作为一个司机分类;

或者,所述分类单元对各司机的决策向量中的拒单因素向量进行聚类,将聚类得到的每一簇中的拒单因素向量对应的司机作为一个司机分类。

在线叫车服务平台中的司机分类方法和装置

【技术领域】

[0001] 本发明涉及互联网技术,特别涉及在线叫车服务平台中的司机分类方法和装置。

【背景技术】

[0002] 在线叫车服务中,司机接单行为是决定叫车成功与否的关键。由于在接单过程中考虑的因素不一样,不同的司机可能具有不同的接单行为。如何将具有类似行为的司机划分到同一类,对于分析司机的接单行为具有重要意义。

[0003] 现有方式中,主要是通过比较单一的少数指标来对司机进行分类,如接单率、在线时长等指标,分类结果很不准确。

【发明内容】

[0004] 本发明提供了在线叫车服务平台中的司机分类方法和装置,能够提高分类结果的准确性。

[0005] 具体技术方案如下:

[0006] 一种在线叫车服务平台中的司机分类方法,包括:

[0007] 获取训练样本,每个训练样本中包括:订单信息、司机状态信息以及司机接单与否信息,并根据所述训练样本训练得到司机接单行为预测模型;

[0008] 针对每个司机,分别根据派送给所述司机的订单的订单信息、所述司机的司机状态信息以及所述司机接单行为预测模型,将影响所述司机的接单行为的因素量化为所述司机的决策向量;

[0009] 根据获取到的各司机的决策向量对各司机进行分类。

[0010] 一种在线叫车服务平台中的司机分类装置,包括:模型训练单元、确定单元以及分类单元;

[0011] 所述模型训练单元,用于获取训练样本,每个训练样本中包括:订单信息、司机状态信息以及司机接单与否信息,并根据所述训练样本训练得到司机接单行为预测模型,将所述司机接单行为预测模型发送给所述确定单元;

[0012] 所述确定单元,用于针对每个司机,分别根据派送给所述司机的订单的订单信息、所述司机的司机状态信息以及所述司机接单行为预测模型,将影响所述司机的接单行为的因素量化为所述司机的决策向量,并将所述决策向量发送给所述分类单元;

[0013] 所述分类单元,用于根据获取到的各司机的决策向量对各司机进行分类。

[0014] 基于上述介绍可以看出,采用本发明所述方案,可基于由订单信息、司机状态信息以及司机接单与否信息组成的训练样本训练得到司机接单行为预测模型,进而可通过司机接单行为预测模型将影响司机的接单行为的因素量化为司机的决策向量,并可根据各司机的决策向量对各司机进行分类,从而相比于现有技术中通过比较单一的少数指标来对司机进行分类的方式,较好地提高了分类结果的准确性。

【附图说明】

[0015] 图1为本发明所述在线叫车服务平台中的司机分类方法实施例的流程图。

[0016] 图2为本发明所述确定出任一司机a的决策向量的方法实施例的流程图。

[0017] 图3为本发明所述在线叫车服务平台中的司机分类装置实施例的组成结构示意图。

【具体实施方式】

[0018] 为了使本发明的技术方案更加清楚、明白，以下参照附图并举实施例，对本发明所述方案作进一步地详细说明。

[0019] 实施例一

[0020] 图1为本发明所述在线叫车服务平台中的司机分类方法实施例的流程图，如图1所示，包括以下具体实现方式：

[0021] 在11中，获取训练样本，每个训练样本中包括：订单信息、司机状态信息以及司机接单与否信息；

[0022] 在12中，根据训练样本训练得到司机接单行为预测模型；

[0023] 在13中，针对每个司机，分别根据派送给该司机的订单的订单信息、该司机的司机状态信息以及司机接单行为预测模型，将影响该司机的接单行为的因素量化为该司机的决策向量；

[0024] 在14中，根据获取到的各司机的决策向量对各司机进行分类。

[0025] 以下分别对上述各部分内容的具体实现进行详细说明。

[0026] 一) 获取训练样本

[0027] 为实现本发明所述方案，需要根据各司机的历史接单行为等建立司机接单行为预测模型，进而利用司机接单行为预测模型，将各种影响司机接单行为的因素量化为司机的决策向量。

[0028] 而为了得到司机接单行为预测模型，需要首先获取训练样本，进而通过训练样本训练得到司机接单行为预测模型。

[0029] 每个训练样本中可包括：订单信息、司机状态信息以及司机接单与否信息。其中，订单信息为过往派送的任一订单的订单信息，司机状态信息为被派送该订单的任一司机在该订单发生时的司机状态信息。

[0030] 比如，对于用户提交的一个订单，可获取该订单的订单信息，假设该订单被派送给司机a，那么该订单的订单信息、司机a的司机状态信息以及司机a是否接单信息即可组成一个训练样本，如果司机a接单，可记为1，否则，可记为0。

[0031] 订单信息中可包括：订单的起点位置和终点位置、起点位置和终点位置是否为商圈、起点位置和终点位置是否为交通枢纽、所在城市、订单出发时间、起点位置和终点位置之间的距离、订单预计价格、订单预计价格与所在城市历史实际成单平均价格比、订单预计行驶时间、订单预计行驶速度与所在城市历史实际成单平均行驶速度比、是否穿越拥堵区域、提交订单的用户性别等。

[0032] 其中，订单的起点位置和终点位置需要转化成区域编码，转换成区域编码的方式

有多种,比如可以采用GeoHash方式,也可以是按照经纬度划分为矩形或者六边形等形状的区域,然后分别对不同的区域赋予不同的id或编号。

[0033] 订单的出发时间可以包含多个维度,如可包括:是否上午、是否下午、是否晚上、是否后半夜、星期、小时、是否周末、是否为上下班高峰等。

[0034] 司机状态信息可包括:位置相关信息以及历史信息。

[0035] 其中,位置相关信息可包括:当前移动速度、当前移动状态持续时间、当前位置(也需要转化成区域编码)、接人行驶距离、预期接人行驶时间、预期接人平均行驶速度等。

[0036] 历史信息可包括:司机过去M天推单接单比、司机过去M天平均在线时长与所在城市所有司机过去M天平均在线时长比、司机过去M天在线时长的4分位数与所在城市所有司机过去M天在线时长的4分位数均值比、司机过去M天平均推单次数与所在城市所有司机过去M天平均推单次数比、司机过去M天平均接单次数与所在城市所有司机过去M天平均接单次数比、司机过去M天轨迹中的移动状态时间占比、司机过去M天轨迹中的移动状态时间占比与所在城市所有司机过去M天轨迹中的移动状态时间占比均值的比值等。

[0037] 上述订单信息和司机状态信息中具体包括哪些信息可根据实际需要而定,不限于以上所示。

[0038] 按照上述方式,可得到大量的训练样本。

[0039] 二) 司机接单行为预测模型

[0040] 在得到足够数量的训练样本之后,即可根据这些训练样本进行司机接单行为预测模型的训练。

[0041] 所述司机接单行为预测模型为决策树模型,比如,可为随机森林(Random Forest)模型或梯度提升决策树(GBDT,Gradient Boosting Decision Tree)模型等。

[0042] 如何训练得到司机接单行为预测模型为现有技术。

[0043] 后续,针对用户提交的任一订单,可以利用司机接单行为预测模型,预测出不同司机的接单概率,并可将接单概率与预定阈值进行比较,大于阈值,则认为司机会接单,否则拒单。

[0044] 三) 司机的决策向量

[0045] Random Forest模型和GBDT模型都是由多棵决策树构成的,最终的结果由多棵决策树共同决策。

[0046] 针对上述决策树模型,图2为本发明所述确定出任一司机a的决策向量的方法实施例的流程图,如图2所示,包括以下具体实现方式。

[0047] 在21中,针对决策树模型中的每棵决策树,分别按照22~23所示方式进行处理。

[0048] 在22中,对于最近预定时长内派送给司机a的每个订单,分别根据该订单的订单信息和司机a的司机状态信息,确定出该订单的因素向量。

[0049] 所述预定时长的具体取值可根据实际需要而定,比如,最近一个月。

[0050] 在决策树模型的决策树的建立过程中,每次决策树在某个节点上的分裂都需要搜索在所有的属性的基础上分裂带来的纯净度收益,所述收益可以选用现有决策树模型中常用的标准,如gini系数、信息增益或信息增益比等,记录在建树过程中计算的每个非叶子节点上分裂时各属性的纯净度收益。

[0051] 对于最近预定时长内派送给司机a的每个订单如订单o,可首先确定出根据订单o

的订单信息和司机a的司机状态信息,在利用决策树模型对司机a的接单行为进行决策的过程中,在该决策树上走过的路径p。

[0052] 可以用路径p上的各属性的收益来反映司机接单行为决策过程考虑的因素,为此,可分别获取路径p上的每个非叶子节点的属性收益向量 $\text{vec}(t)$ 。

[0053] $\text{vec}(t) = (\text{Gain}(1,t), \text{Gain}(2,t), \dots, \text{Gain}(i,t), \dots, \text{Gain}(N,t))$;

[0054] 其中,t表示路径p中的任一非叶子节点;

[0055] N表示属性个数,属性个数为订单o的订单信息中包括的信息数与司机a的司机状态信息中包括的信息数之和,订单信息中包括的每条信息以及司机状态信息中包括的每条信息分别为一个属性;

[0056] $\text{Gain}(i,t)$ 表示非叶子节点t上分裂时第i个属性的纯净度收益。

[0057] 如一)中所述,订单信息中可包括:订单出发时间、订单预计价格等,司机状态信息中可包括:当前移动速度、司机过去M天推单接单比等,那么,订单出发时间即为一个属性,当前移动速度也是一个属性。

[0058] 之后,可分别将路径p上的各非叶子节点的属性收益向量相加,用相加之和除以路径p上的非叶子节点数,从而得到订单o的因素向量 $e(o,p)$ 。

[0059] 即有: $e(o,p) = \sum_{t \in p} \text{vec}(t) / \text{Length}(p)$; (1)

[0060] 其中, $\text{Length}(p)$ 表示路径p上的非叶子节点数,如路径p上总共有3个非叶子节点,则 $\text{Length}(p)$ 的取值为3。

[0061] 两个向量的和为两个向量中相同位置的元素的和组成的向量。

[0062] 在23中,根据最近预定时长内的各订单的因素向量以及各订单的司机接单与否信息,确定出该决策树对应的决策向量。

[0063] 由于司机的行为可分为接单和拒单,因此,可分别计算最近预定时长内司机a接受的每个订单的因素向量与对应的加权系数的乘积,将各乘积相加,用相加之和除以最近预定时长内司机a接受的订单数,得到接单因素向量;并分别计算最近预定时长内司机a拒接的每个订单的因素向量与对应的加权系数的乘积,将各乘积相加,用相加之和除以最近预定时长内司机a拒接的订单数,得到拒单因素向量。

[0064] 即,接单因素向量 $\text{accept}_{\text{vec}} = \sum_{o \in \text{acceptorder}} w_o e(o,p) / \text{number_of_accept_order}$; (2)

[0065] 式2)中,o表示最近预定时长内司机a接受的任一订单, w_o 表示订单o对应的加权系数, $e(o,p)$ 表示订单o的因素向量, $\text{number_of_accept_order}$ 表示最近预定时长内司机a接受的订单数。

[0066] 式2)中,各加权系数的具体取值可根据实际需要而定,比如,可全部取值为1,表示对各订单同等对待,也可以针对司机a接受的每个订单,分别将根据决策树模型预测出的司机a的接单概率作为该订单的因素向量对应的加权系数,即订单 $e(o,p)$ 对应的 w_o 可为针对订单o预测出的司机a的接单概率。

[0067] $\text{accept}_{\text{vec}}$ 向量中的各元素与前面所述的各属性一一对应,某一元素的取值越大,表示该因素在司机接单决策过程中的作用越大。

[0068] 拒单因素向量 $\text{reject}_{\text{vec}} = \sum_{o \in \text{rejectorder}} w_o e(o,p) / \text{number_of_reject_order}$; (3)

[0069] 式3)中,o表示最近预定时长内司机a拒接的任一订单, w_o 表示订单o对应的加权系数, $e(o,p)$ 表示订单o的因素向量, $\text{number_of_reject_order}$ 表示最近预定时长内司机a拒

接的订单数。

[0070] 类似地,式3)中,各加权系数的具体取值可根据实际需要而定,比如,可全部取值为1,表示对各订单同等对待,也可以针对司机a拒接的每个订单,分别将根据决策树模型预测出的司机a的拒单概率作为该订单的因素向量对应的加权系数,如可用1减去接单概率,得到拒单概率。

[0071] $reject_{vec}$ 向量中的各元素与前面所述的各属性一一对应,某一元素的取值越大,表示该因素在司机拒单决策过程中的作用越大。

[0072] 在分别得到接单因素向量和拒单因素向量之后,可利用二者组成决策向量 $Decision_{vec}$,即有: $Decision_{vec} = (accept_{vec}, reject_{vec})$ 。

[0073] 在24中,根据决策树模型中的各决策树对应的决策向量确定出司机a的决策向量。

[0074] 按照22~23所述方式,可分别确定出决策树模型中的每棵决策树对应的决策向量,之后,可将决策树模型中的各决策树对应的决策向量相加,并用相加之和除以决策树模型中的决策树数,得到司机a的决策向量。

[0075] 比如,决策树模型中共包括3棵决策树,针对每棵决策树,分别计算出一个对应于司机a的决策向量,那么司机a的决策向量则为将这3个决策向量相加之后除以3的结果。

[0076] 四) 司机分类

[0077] 按照3)中所述方式,可分别获取到每个司机的决策向量,之后,可根据各司机的决策向量对各司机进行分类。

[0078] 分类方式可为以下之一:

[0079] 方式一

[0080] 对各司机的决策向量进行聚类,将聚类得到的每一簇(即每一个聚类结果)中的决策向量对应的司机作为一个司机分类;

[0081] 方式二

[0082] 对各司机的决策向量中的接单因素向量进行聚类,将聚类得到的每一簇中的接单因素向量对应的司机作为一个司机分类;

[0083] 方式三

[0084] 对各司机的决策向量中的拒单因素向量进行聚类,将聚类得到的每一簇中的拒单因素向量对应的司机作为一个司机分类。

[0085] 如果希望综合接单行为和拒单行为对各司机进行分类,则可采用方式一,如果只是希望根据接单行为来对各司机进行分类,则可采用方式二,如果只是希望根据拒单行为来对各司机进行分类,则可采用方式三。

[0086] 所采用的聚类算法可以是基于密度的聚类算法,如具有噪声的基于密度的聚类算法(DBSCAN,Density-Based Spatial Clustering of Applications with Noise),也可以是分层聚类算法,如Ward算法,还可以是基于距离的聚类算法,如K-Means、Mean-shift算法。

[0087] 聚类过程中采用的距离可以为Minkowsky距离中的一种,如曼哈顿距离或欧式距离等。

[0088] 假设共存在10个司机需要进行分类,分别为司机1~司机10,每个司机分别对应一个决策向量,通过聚类,将这10个决策向量分为3个簇,其中一个簇中包括3个决策向量,分

别对应司机1~司机3,那么司机1~司机3则被分为一类,另一个簇中也包括3个决策向量,分别对应司机4~司机6,那么司机4~司机6则被分为一类,剩下的一个簇中包括4个决策向量,分别对应司机7~司机10,那么司机7~司机10则被分为一类。

[0089] 在实际应用中,当采集到足够的训练样本并训练得到决策树模型之后,可利用决策树模型来收集各司机的决策向量信息,如收集时长可设为一个月,进而可根据收集结果对各司机进行分类。

[0090] 以上是关于方法实施例的介绍,以下通过装置实施例,对本发明所述方案进行进一步说明。

[0091] 实施例二

[0092] 图3为本发明所述在线叫车服务平台中的司机分类装置实施例的组成结构示意图,如图3所示,包括:模型训练单元31、确定单元32以及分类单元33。

[0093] 模型训练单元31,用于获取训练样本,每个训练样本中包括:订单信息、司机状态信息以及司机接单与否信息,并根据训练样本训练得到司机接单行为预测模型,将司机接单行为预测模型发送给确定单元32。

[0094] 确定单元32,用于针对每个司机,分别根据派送给该司机的订单的订单信息、该司机的司机状态信息以及司机接单行为预测模型,将影响该司机的接单行为的因素量化为该司机的决策向量,并将决策向量发送给分类单元33。

[0095] 分类单元33,用于根据获取到的各司机的决策向量对各司机进行分类。

[0096] 为实现本发明所述方案,需要根据各司机的历史接单行为等建立司机接单行为预测模型,进而利用司机接单行为预测模型,将各种影响司机接单行为的因素量化为司机的决策向量。

[0097] 而为了得到司机接单行为预测模型,需要首先获取训练样本,进而通过训练样本训练得到司机接单行为预测模型。

[0098] 每个训练样本中可包括:订单信息、司机状态信息以及司机接单与否信息。其中,订单信息为过往派送的任一订单的订单信息,司机状态信息为被派送该订单的任一司机在该订单发生时的司机状态信息。

[0099] 比如,对于用户提交的一个订单,可获取该订单的订单信息,假设该订单被派送给司机a,那么该订单的订单信息、司机a的司机状态信息以及司机a是否接单信息即可组成一个训练样本,如果司机a接单,可记为1,否则,可记为0。

[0100] 订单信息中可包括:订单的起点位置和终点位置、起点位置和终点位置是否为商圈、起点位置和终点位置是否为交通枢纽、所在城市、订单出发时间、起点位置和终点位置之间的距离、订单预计价格、订单预计价格与所在城市历史实际成单平均价格比、订单预计行驶时间、订单预计行驶速度与所在城市历史实际成单平均行驶速度比、是否穿越拥堵区域、提交订单的用户性别等。

[0101] 其中,订单的起点位置和终点位置需要转化成区域编码,转换成区域编码的方式有多种,比如可以采用GeoHash方式,也可以是按照经纬度划分为矩形或者六边形等形状的区块,然后分别对不同的区块赋予不同的id或编号。

[0102] 订单的出发时间可以包含多个维度,如可包括:是否上午、是否下午、是否晚上、是否后半夜、星期、小时、是否周末、是否为上下班高峰等。

[0103] 司机状态信息可包括：位置相关信息以及历史信息。

[0104] 其中，位置相关信息可包括：当前移动速度、当前移动状态持续时间、当前位置（也需要转化成区域编码）、接人行驶距离、预期接人行驶时间、预期接人平均行驶速度等。

[0105] 历史信息可包括：司机过去M天推单接单比、司机过去M天平均在线时长与所在城市所有司机过去M天平均在线时长比、司机过去M天在线时长的4分位数与所在城市所有司机过去M天在线时长的4分位数均值比、司机过去M天平均推单次数与所在城市所有司机过去M天平均推单次数比、司机过去M天平均接单次数与所在城市所有司机过去M天平均接单次数比、司机过去M天轨迹中的移动状态时间占比、司机过去M天轨迹中的移动状态时间占比与所在城市所有司机过去M天轨迹中的移动状态时间占比均值的比值等。

[0106] 上述订单信息和司机状态信息中具体包括哪些信息可根据实际需要而定，不限于以上所示。

[0107] 按照上述方式，模型训练单元31可得到大量的训练样本。

[0108] 之后，模型训练单元31可根据这些训练样本训练得到司机接单行为预测模型。

[0109] 司机接单行为预测模型为决策树模型，比如，可为Random Forest模型或GBDT模型等。

[0110] 如图3所示，确定单元32中可具体包括：第一处理子单元321和第二处理子单元322。

[0111] 第一处理子单元321，用于针对每个司机，分别按照以下方式获取决策树模型中的每棵决策树对应的决策向量：对于最近预定时长内派送给该司机的每个订单，分别根据该订单的订单信息和该司机的司机状态信息，确定出该订单的因素向量，根据最近预定时长内的各订单的因素向量以及各订单的接单与否信息，确定出该决策树对应的决策向量；分别将各决策树对应的决策向量发送给第二处理子单元322。

[0112] 第二处理子单元322，用于针对每个司机，分别根据获取到的该司机对应的、决策树模型中的各决策树对应的决策向量，确定出该司机的决策向量，并发送给分类单元33。

[0113] 具体地，第一处理子单元321可确定出根据订单信息和司机状态信息，利用决策树模型对该司机的接单行为进行决策的过程中，在该决策树上走过的路径；分别获取该路径上的每个非叶子节点的属性收益向量 $\text{vec}(t)$ ， $\text{vec}(t) = (\text{Gain}(1, t), \text{Gain}(2, t), \dots, \text{Gain}(i, t), \dots, \text{Gain}(N, t))$ ，并将该路径上的各非叶子节点的属性收益向量相加，用相加之和除以该路径上的非叶子节点数，得到该订单的因素向量。

[0114] 第一处理子单元321可分别计算最近预定时长内该司机接受的每个订单的因素向量与对应的加权系数的乘积，并将各乘积相加，用相加之和除以最近预定时长内该司机接受的订单数，得到接单因素向量；分别计算最近预定时长内该司机拒接的每个订单的因素向量与对应的加权系数的乘积，并将各乘积相加，用相加之和除以最近预定时长内该司机拒接的订单数，得到拒单因素向量；用接单因素向量和所述拒单因素向量组成该决策树对应的决策向量。

[0115] 其中，在计算接单因素向量时，第一处理子单元321可以将每个加权系数的取值均设置为1，或者，针对该司机接受的每个订单，分别将根据决策树模型预测出来的该司机的接单概率作为该订单的因素向量对应的加权系数。

[0116] 类似地，在计算拒单因素向量时，第一处理子单元321可将每个加权系数的取值均

设置为1,或者,针对该司机拒接的每个订单,分别将根据决策树模型预测出来的该司机的拒单概率作为该订单的因素向量对应的加权系数。

[0117] 这样,针对每个司机,在分别得到每棵决策树对应的决策向量之后,可由第二处理子单元322将决策树模型中的各决策树对应的决策向量相加,并用相加之和除以决策树模型中的决策树数,从而得到该司机的决策向量。

[0118] 在分别获取到每个司机的决策向量之后,可由分类单元33根据各司机的决策向量对各司机进行分类。

[0119] 分类方式可为以下之一:

[0120] 方式一

[0121] 分类单元33对各司机的决策向量进行聚类,将聚类得到的每一簇中的决策向量对应的司机作为一个司机分类;

[0122] 方式二

[0123] 分类单元33对各司机的决策向量中的接单因素向量进行聚类,将聚类得到的每一簇中的接单因素向量对应的司机作为一个司机分类;

[0124] 方式三

[0125] 分类单元33对各司机的决策向量中的拒单因素向量进行聚类,将聚类得到的每一簇中的拒单因素向量对应的司机作为一个司机分类。

[0126] 如果希望综合接单行为和拒单行为对各司机进行分类,则可采用方式一,如果只是希望根据接单行为来对各司机进行分类,则可采用方式二,如果只是希望根据拒单行为来对各司机进行分类,则可采用方式三。

[0127] 图3所示装置实施例的具体工作流程请参照前述方法实施例中的相应说明,此处不再赘述。

[0128] 总之,采用本发明所述方案,可基于由订单信息、司机状态信息以及司机接单与否信息组成的训练样本训练得到司机接单行为预测模型,进而可通过司机接单行为预测模型将影响司机的接单行为的因素量化为司机的决策向量,并可根据各司机的决策向量对各司机进行分类,从而相比于现有技术中通过比较单一的少数指标来对司机进行分类的方式,较好地提高了分类结果的准确性;而且,本发明所述方案可适用于各种在线叫车服务平台,具有广泛适用性。

[0129] 在本发明所提供的几个实施例中,应该理解到,所揭露的装置和方法,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。

[0130] 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。

[0131] 另外,在本发明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。

[0132] 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,服务器,或者网络设备等)或处理器(processor)执行本发明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(ROM, Read-Only Memory)、随机存取存储器(RAM, Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0133] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,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,均应包含在本发明保护的范围之内。

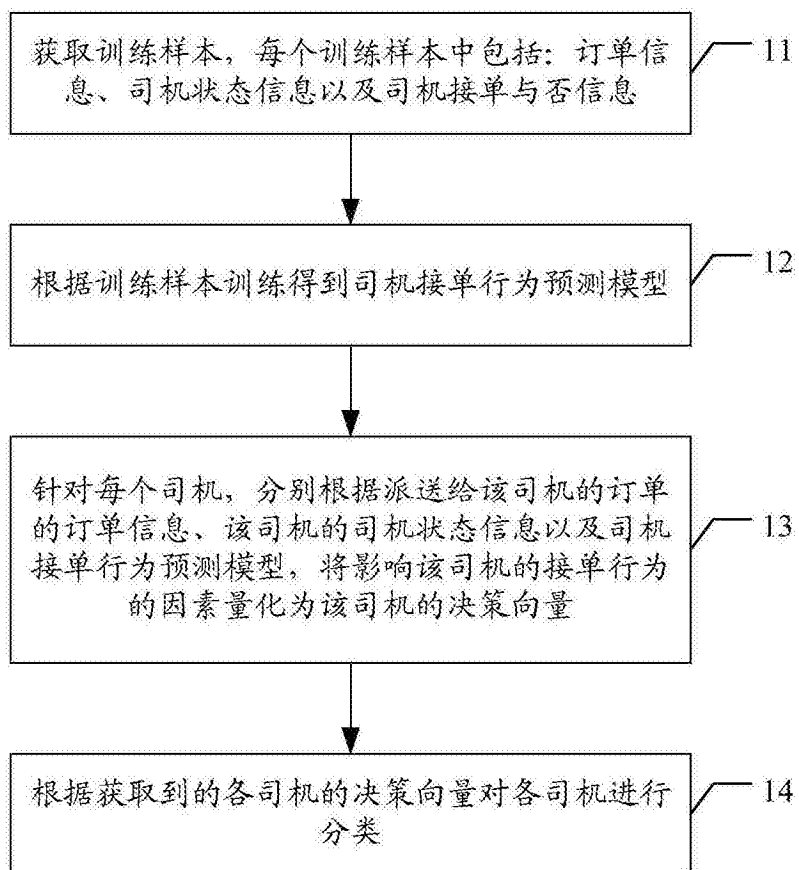


图1

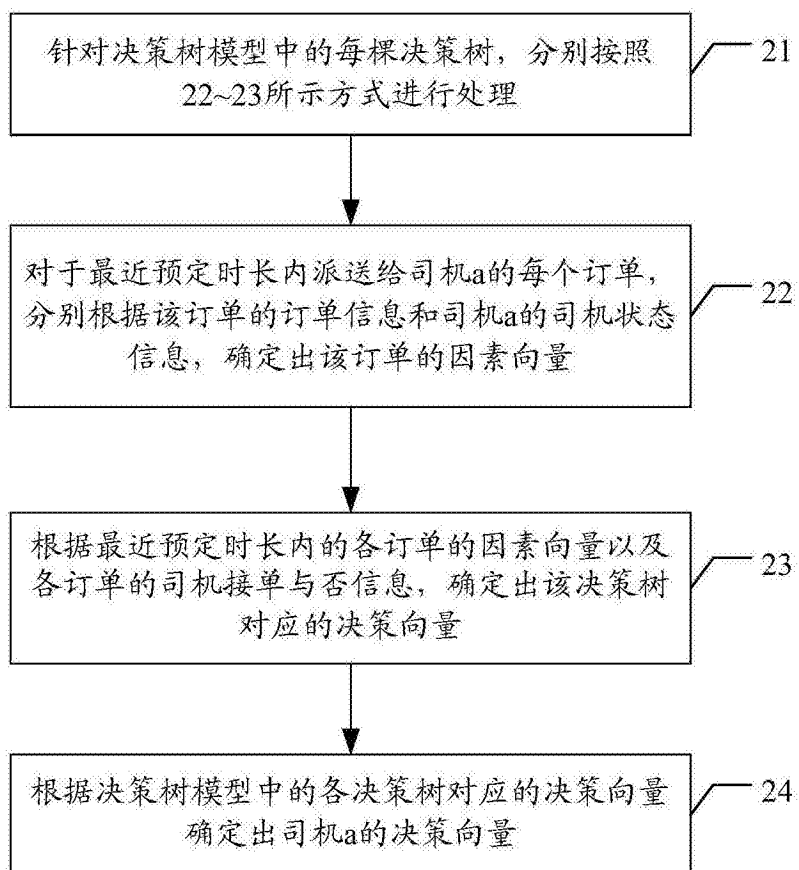


图2

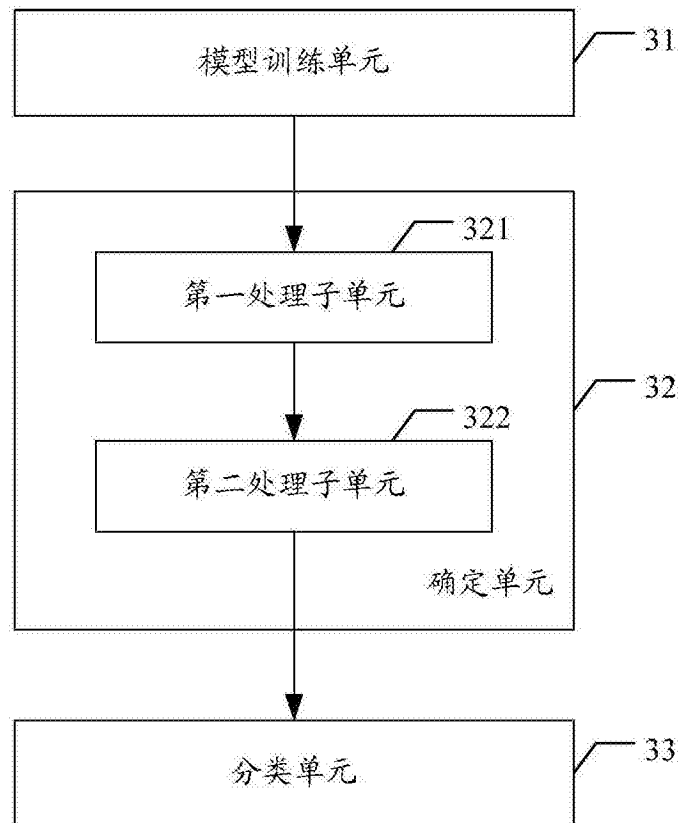


图3