



## (12) 发明专利

(10) 授权公告号 CN 113821703 B

(45) 授权公告日 2023.12.08

(21) 申请号 202010557455.9

(22) 申请日 2020.06.18

(65) 同一申请的已公布的文献号

申请公布号 CN 113821703 A

(43) 申请公布日 2021.12.21

(73) 专利权人 广州汽车集团股份有限公司

    地址 510030 广东省广州市越秀区东风中  
    路448—458号成悦大厦23楼

(72) 发明人 李林 谢亚果 霍广 郭和攀

(74) 专利代理机构 深圳汇智容达专利商标事务  
    所(普通合伙) 44238

专利代理人 徐文城

(51) Int.Cl.

G06F 16/951 (2019.01)

G06F 16/9535 (2019.01)

G06F 18/23213 (2023.01)

G06Q 30/0201 (2023.01)

G06Q 30/0601 (2023.01)

(54) 发明名称

一种车联网用户画像生成方法及其系统

(57) 摘要

本发明涉及一种车联网用户画像生成方法及其系统，所述方法包括：根据用户身份信息获取车联网中多个数据源的原始用户数据，每一数据源的原始用户数据包括用户的基础属性数据、消费行为数据、驾驶行为数据、娱乐行为数据中的一种或多种；按照预设处理规则对所述多个数据源的原始用户数据进行预处理得到待标记用户数据；按照预设标签规则对所述待标记用户数据进行标签标记得到用户标签数据，所述用户标签数据包括多个用户数据及其标签；对所述用户标签数据进行聚类分析得到多个用户数据集合；根据所述多个用户数据集合计算每一标签的比重和稀缺度，并根据所述每一标签的比重和稀缺度对用户进行多维度描述，以生成用户画像。相对于现有技术，本发明能够生成包括用户娱乐行

CN 113821703 B

(56) 对比文件

CN 106504099 A, 2017.03.15

CN 109784367 A, 2019.05.21

CN 109933699 A, 2019.06.25

CN 110674144 A, 2020.01.10

US 2017004557 A1, 2017.01.05

US 2019102652 A1, 2019.04.04

CN 110111136 A, 2019.08.09

CN 111126880 A, 2020.05.08

CN 109978630 A, 2019.07.05

CN 111191125 A, 2020.05.22

CN 106504534 A, 2017.03.15

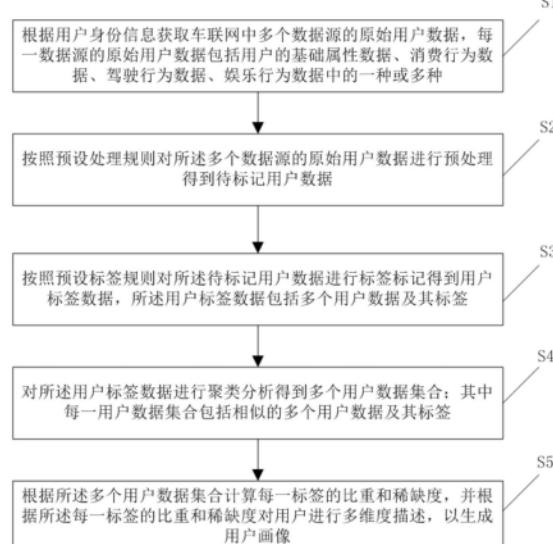
CN 107688606 A, 2018.02.13

程兴乾.基于车联网的用户驾驶行为统计分析研究.《中国优秀硕士学位论文全文数据库 工程科技Ⅱ辑》.2020,C034-986. (续)

审查员 李若晨

权利要求书3页 说明书8页 附图2页

为、驾驶行为、消费行为等维度更为丰富的用户画像。



[转续页]

[接上页]

(56) 对比文件

许甜华.用户画像技术在电子商务系统中的研究与应用.《中国优秀硕士学位论文全文数据  
库 信息科技辑》.2019, I138-1529.

Zhong, Chunlin等.Research on Electricity Consumption Behavior of Electric Power Users Based on Tag Technology and Clustering Algorithm.《2018 5TH INTERNATIONAL CONFERENCE ON

INFORMATION SCIENCE AND CONTROL ENGINEERING (ICISCE 2018)》.2019, 第459-462 页.

Weijie Wang等.Research on Awareness Method of Cloud User Abnormal Behavior Based on Log Audit.《2018 IEEE 4th International Conference on Computer and Communications (ICCC) . Proceedings》.2019, 第1944-50页.

1.一种车联网用户画像生成方法,其特征在于,包括:

根据用户身份信息获取车联网中多个数据源的原始用户数据,每一数据源的原始用户数据包括用户的基础属性数据、消费行为数据、驾驶行为数据、娱乐行为数据中的一种或多种;

按照预设处理规则对所述多个数据源的原始用户数据进行预处理得到待标记用户数据;

按照预设标签规则对所述待标记用户数据进行标签标记得到用户标签数据,所述用户标签数据包括多个用户数据及其标签;

对所述用户标签数据进行聚类分析得到多个用户数据集合;其中每一用户数据集合包括相似的多个用户数据及其标签;

根据所述用户标签数据构建标签树模型,在所述标签树模型中,若某一用户数据集合包括某一标签,则该用户数据集合作为一个尾节点被连接至对应的上层标签节点;

根据所述多个用户数据集合以及所述标签树模型计算每一标签的比重和稀缺度,并根据所述每一标签的比重和稀缺度对用户进行多维度描述,以生成用户画像;

其中,所述根据所述多个用户数据集合以及所述标签树模型对用户进行多维度描述,包括:

根据公式 $TF_i = W_i / N_1$ 计算每一用户标签的比重;其中, $TF_i$ 为用户标签*i*的比重, $W_i$ 为用户标签*i*所对应的用户数据集合中用户标签*i*的个数, $N_1$ 为多个用户数据集合的标签总数;

根据公式 $IDF_i = M_i / N_2$ 计算每一用户标签的稀缺度;其中, $IDF_i$ 为用户标签*i*的稀缺度, $M_i$ 为用户数据集合的个数, $N_2$ 为包含用户标签*i*所对应的标签子节点所连接的尾节点个数;

根据每一用户标签的比重和稀缺度分别对每一用户标签进行多维度描述,并可视化描述内容,以生成用户画像。

2.根据权利要求1所述的车联网用户画像生成方法,其特征在于,所述基础属性数据包括用户的性别信息、年龄信息、地域信息、学历信息、收入信息中的一种或多种;所述消费行为数据包括购物信息、停车信息、加油信息、充电信息、流量充值信息、积分兑换信息中的一种或多种;所述驾驶行为数据包括油耗信息、速度信息、里程信息、电耗信息、驾驶行为评分信息、保险信息中的一种或多种;所述娱乐行为数据包括音乐偏好信息、电台偏好信息、视频偏好信息、游戏偏好信息中的一种或多种。

3.如权利要求1所述的车联网用户画像生成方法,其特征在于,所述对所述多个用户数据进行聚类分析得到多个用户数据集合,包括:

S41、随机确定一个k值;

S42、从所述多个用户数据中随机选择k个用户数据作为质心;

S43、对所述多个用户数据中的每一个用户数据,计算其与每一个质心的距离,并将其划分至与其距离最小的质心所属的用户数据集合,以得到k个用户数据集合;

S44、重新计算所述k个用户数据集合的质心;

S45、如果重新计算出来的用户数据集合质心和原来的用户数据集合质心之间的距离小于预设阈值,则结束聚类;如果重新计算出来的用户数据集合质心和原来的用户数据集合质心之间的距离大于等于预设阈值,令 $k=k+1$ ,并返回步骤S42继续聚类。

4.如权利要求3所述的车联网用户画像生成方法,其特征在于,所述标签树模型的首节

点包括基础属性节点、消费行为节点、驾驶行为节点以及娱乐行为节点；

其中，所述基础属性节点的子节点包括性别标签节点、年龄标签节点、地域标签节点、学历标签节点、收入标签节点中的一个或多个，所述消费行为节点的子节点包括购物标签节点、停车标签节点、加油标签节点、充电标签节点、流量充值标签节点、积分兑换标签节点中的一种或多种；所述驾驶行为节点的子节点包括油耗标签节点、速度标签节点、里程标签节点、电耗标签节点、驾驶行为评分标签节点、保险标签节点中的一种或多种；所述娱乐行为节点的子节点包括音乐偏好标签节点、电台偏好标签节点、视频偏好标签节点、游戏偏好标签节点中的一种或多种；每一子节点连接多个尾节点。

5. 一种车联网用户画像生成系统，其特征在于，包括：

原始数据获取单元，用于根据用户身份信息获取车联网中多个数据源的原始用户数据，每一数据源的原始用户数据包括用户的基础属性数据、消费行为数据、驾驶行为数据、娱乐行为数据中的一种或多种；

预处理单元，用于按照预设处理规则对所述多个数据源的原始用户数据进行预处理得到待标记用户数据；

数据标记单元，用于按照预设标签规则对所述待标记用户数据进行标签标记得到用户标签数据，所述用户标签数据包括多个用户数据及其标签；

聚类分析单元，用于对所述用户标签数据进行聚类分析得到多个用户数据集合；其中每一用户数据集合包括相似的多个用户数据及其标签；

标签树构建单元，用于根据所述用户标签数据构建标签树模型，在所述标签树模型中，若某一用户数据集合包括某一标签，则该用户数据集合作为一个尾节点被连接至对应的上层标签节点；以及

画面描述单元，用于根据所述多个用户数据集合以及所述标签树模型计算每一标签的比重和稀缺度，并根据所述每一标签的比重和稀缺度对用户进行多维度描述，以生成用户画像；

其中，所述画面描述单元，具体用于：

根据公式 $TF_i = W_i / N1$ 计算每一用户标签的比重；其中， $TF_i$ 为用户标签*i*的比重， $W_i$ 为用户标签*i*所对应的用户数据集合中用户标签*i*的个数， $N1$ 为多个用户数据集合的标签总数；

根据公式 $IDF_i = M_i / N2$ 计算每一用户标签的稀缺度；其中， $IDF_i$ 为用户标签*i*的稀缺度， $M_i$ 为用户数据集合的个数， $N2$ 为包含用户标签*i*所对应的标签子节点所连接的尾节点个数；

根据每一用户标签的比重和稀缺度分别对每一用户标签进行多维度描述，并可视化描述内容，以生成用户画像。

6. 如权利要求5所述的车联网用户画像生成系统，其特征在于，所述基础属性数据包括用户的性别信息、年龄信息、地域信息、学历信息、收入信息中的一种或多种；所述消费行为数据包括购物信息、停车信息、加油信息、充电信息、流量充值信息、积分兑换信息中的一种或多种；所述驾驶行为数据包括油耗信息、速度信息、里程信息、电耗信息、驾驶行为评分信息、保险信息中的一种或多种；所述娱乐行为数据包括音乐偏好信息、电台偏好信息、视频偏好信息、游戏偏好信息中的一种或多种。

7. 如权利要求5所述的车联网用户画像生成系统，其特征在于，

所述标签树模型的首节点包括基础属性节点、消费行为节点、驾驶行为节点以及娱乐

行为节点；所述基础属性节点的子节点包括性别标签节点、年龄标签节点、地域标签节点、学历标签节点、收入标签节点中的一个或多个，所述消费行为节点的子节点包括购物标签节点、停车标签节点、加油标签节点、充电标签节点、流量充值标签节点、积分兑换标签节点中的一种或多种；所述驾驶行为节点的子节点包括油耗标签节点、速度标签节点、里程标签节点、电耗标签节点、驾驶行为评分标签节点、保险标签节点中的一种或多种；所述娱乐行为节点的子节点包括音乐偏好标签节点、电台偏好标签节点、视频偏好标签节点、游戏偏好标签节点中的一种或多种；每一子节点连接多个尾节点。

## 一种车联网用户画像生成方法及其系统

### 技术领域

[0001] 本发明涉及用户画像构建技术领域,具体涉及一种车联网用户画像生成方法及其系统。

### 背景技术

[0002] 随着车联网的发展以及新能源车上市的标准要求,越来越多的车辆配备车联网功能。基于车联网,车辆用户可以定位、远程控制车辆,甚至可以在车机访问互联网。车厂也可以通过定义信号接收模块采集车辆各个ECU节点的信号,这些信号可以用于生成用户画像,包括用户驾驶、用车的行为。此外,通过对外拓展第三方生态应用数据,可以获取用户的互联网行为。用户画像一方面给车厂对用户用车间况、使用习惯有一个概览,作为下一代的车型开发设计的参考。另一方面,用户画像也可以方便车厂有针对性地提供精准营销服务,提高用户满意度和粘性。

[0003] 当前技术所生成的车联网用户画像缺少了用户的互联网行为数据,例如购物行为、消费行为、娱乐行为、社交行为等,此外,用户的驾驶行为相关数据也没有被利用,因此,亟待提出一种能够生成维度更为丰富的用户画像的方法。

### 发明内容

[0004] 本发明的目的在于提出一种车联网用户画像生成方法及其系统,能够生成包括用户的基础属性、娱乐行为、驾驶行为、消费行为等维度更为丰富的用户画像。

[0005] 为实现上述目的,本发明实施例提出一种车联网用户画像生成方法,包括:

[0006] 根据用户身份信息获取车联网中多个数据源的原始用户数据,每一数据源的原始用户数据包括用户的基础属性数据、消费行为数据、驾驶行为数据、娱乐行为数据中的一种或多种;

[0007] 按照预设处理规则对所述多个数据源的原始用户数据进行预处理得到待标记用户数据;

[0008] 按照预设标签规则对所述待标记用户数据进行标签标记得到用户标签数据,所述用户标签数据包括多个用户数据及其标签;

[0009] 对所述用户标签数据进行聚类分析得到多个用户数据集合;其中每一用户数据集合包括相似的多个用户数据及其标签;

[0010] 根据所述多个用户数据集合计算每一标签的比重和稀缺度,并根据所述每一标签的比重和稀缺度对用户进行多维度描述,以生成用户画像。

[0011] 优选地,所述基础属性数据包括用户的性别信息、年龄信息、地域信息、学历信息、收入信息中的一种或多种;所述消费行为数据包括购物信息、停车信息、加油信息、充电信息、流量充值信息、积分兑换信息中的一种或多种;所述驾驶行为数据包括油耗信息、速度信息、里程信息、电耗信息、驾驶行为评分信息、保险信息中的一种或多种;所述娱乐行为数据包括音乐偏好信息、电台偏好信息、视频偏好信息、游戏偏好信息中的一种或多种。

- [0012] 优选地,所述对所述多个用户数据进行聚类分析得到多个用户数据集合,包括:
- [0013] S41、随机确定一个k值;
- [0014] S42、从所述多个用户数据中随机选择k个用户数据作为质心;
- [0015] S43、对所述多个用户数据中的每一个用户数据,计算其与每一个质心的距离,并将其划分至与其距离最小的质心所属的用户数据集合,以得到k个用户数据集合;
- [0016] S44、重新计算所述k个用户数据集合的质心;
- [0017] S45、如果重新计算出来的用户数据集合质心和原来的用户数据集合质心之间的距离小于预设阈值,则结束聚类;如果重新计算出来的用户数据集合质心和原来的用户数据集合质心之间的距离大于等于预设阈值,令 $k=k+1$ ,并返回步骤S42继续聚类。
- [0018] 优选地,所述根据所述多个用户数据集合计算每一标签的比重和稀缺度,并根据所述每一标签的比重和稀缺度对用户进行多维度描述,以生成用户画像,包括:
- [0019] 根据公式 $TF_i = W_i/N_1$ 计算每一用户标签的比重;其中, $TF_i$ 为用户标签*i*的比重, $W_i$ 为用户标签*i*所对应的用户数据集合中用户标签*i*的个数, $N_1$ 为多个用户数据集合的标签总数;
- [0020] 根据公式 $IDF_i = M_i/N_2$ 计算每一用户标签的稀缺度;其中, $IDF_i$ 为用户标签*i*的稀缺度, $M_i$ 为用户数据集合的个数, $N_2$ 为包含用户标签*i*的用户数据集合的个数;
- [0021] 根据每一用户标签的比重和稀缺度分别对每一用户标签进行多维度描述,并可视化描述内容,以生成用户画像。
- [0022] 优选地,所述方法还包括:根据所述用户标签数据构建标签树模型,在所述标签树模型中,若某一用户数据集合包括某一标签,则该用户数据集合作为一个尾节点被连接至对应的上层标签节点;
- [0023] 其中,所述根据所述多个用户数据集合计算每一标签的比重和稀缺度,并根据所述每一标签的比重和稀缺度对用户进行多维度描述,以生成用户画像,具体为:
- [0024] 根据所述多个用户数据集合以及所述标签树模型计算每一标签的比重和稀缺度,并根据所述每一标签的比重和稀缺度对用户进行多维度描述,以生成用户画像。
- [0025] 优选地,所述标签树模型的首节点包括基础属性节点、消费行为节点、驾驶行为节点以及娱乐行为节点;
- [0026] 其中,所述基础属性节点的子节点包括性别标签节点、年龄标签节点、地域标签节点、学历标签节点、收入标签节点中的一个或多个,所述消费行为节点的子节点包括购物标签节点、停车标签节点、加油标签节点、充电标签节点、流量充值标签节点、积分兑换标签节点中的一种或多种;所述驾驶行为节点的子节点包括油耗标签节点、速度标签节点、里程标签节点、电耗标签节点、驾驶行为评分标签节点、保险标签节点中的一种或多种;所述娱乐行为节点的子节点包括音乐偏好标签节点、电台偏好标签节点、视频偏好标签节点、游戏偏好标签节点中的一种或多种;每一子节点连接多个尾节点。
- [0027] 优选地,所述根据所述多个用户数据集合以及所述标签树模型对用户进行多维度描述,包括:
- [0028] 根据公式 $TF_i = W_i/N_1$ 计算每一用户标签的比重;其中, $TF_i$ 为用户标签*i*的比重, $W_i$ 为用户标签*i*所对应的用户数据集合中用户标签*i*的个数, $N_1$ 为多个用户数据集合的标签总数;

[0029] 根据公式 $IDF_i = M_i / N_2$ 计算每一用户标签的稀缺度；其中， $IDF_i$ 为用户标签*i*的稀缺度， $M_i$ 为用户数据集合的个数， $N_2$ 为包含用户标签*i*所对应的标签子节点所连接的尾节点个数；

[0030] 根据每一用户标签的比重和稀缺度分别对每一用户标签进行多维度描述，并可视化描述内容，以生成用户画像。

[0031] 作为同一发明构思，本发明实施例还一种车联网用户画像生成系统，包括：

[0032] 原始数据获取单元，用于根据用户身份信息获取车联网中多个数据源的原始用户数据，每一数据源的原始用户数据包括用户的基础属性数据、消费行为数据、驾驶行为数据、娱乐行为数据中的一种或多种；

[0033] 预处理单元，用于按照预设处理规则对所述多个数据源的原始用户数据进行预处理得到待标记用户数据；

[0034] 数据标记单元，用于按照预设标签规则对所述待标记用户数据进行标签标记得到用户标签数据，所述用户标签数据包括多个用户数据及其标签；

[0035] 聚类分析单元，用于对所述用户标签数据进行聚类分析得到多个用户数据集合；其中每一用户数据集合包括相似的多个用户数据及其标签；以及

[0036] 画像生成单元，用于根据所述多个用户数据集合计算每一标签的比重和稀缺度，并根据所述每一标签的比重和稀缺度对用户进行多维度描述，以生成用户画像。

[0037] 优选地，所述基础属性数据包括用户的性别信息、年龄信息、地域信息、学历信息、收入信息中的一种或多种；所述消费行为数据包括购物信息、停车信息、加油信息、充电信息、流量充值信息、积分兑换信息中的一种或多种；所述驾驶行为数据包括油耗信息、速度信息、里程信息、电耗信息、驾驶行为评分信息、保险信息中的一种或多种；所述娱乐行为数据包括音乐偏好信息、电台偏好信息、视频偏好信息、游戏偏好信息中的一种或多种。

[0038] 优选地，所述画像生成单元包括：

[0039] 标签树构建单元，用于根据所述用户标签数据构建标签树模型，在所述标签树模型中，若某一用户数据集合包括某一标签，则该用户数据集合作为一个尾节点被连接至对应的上层标签节点；以及

[0040] 画面描述单元，用于根据所述多个用户数据集合以及所述标签树模型计算每一标签的比重和稀缺度，并根据所述每一标签的比重和稀缺度对用户进行多维度描述，以生成用户画像；

[0041] 其中，所述标签树模型的首节点包括基础属性节点、消费行为节点、驾驶行为节点以及娱乐行为节点；所述基础属性节点的子节点包括性别标签节点、年龄标签节点、地域标签节点、学历标签节点、收入标签节点中的一个或多个，所述消费行为节点的子节点包括购物标签节点、停车标签节点、加油标签节点、充电标签节点、流量充值标签节点、积分兑换标签节点中的一种或多种；所述驾驶行为节点的子节点包括油耗标签节点、速度标签节点、里程标签节点、电耗标签节点、驾驶行为评分标签节点、保险标签节点中的一种或多种；所述娱乐行为节点的子节点包括音乐偏好标签节点、电台偏好标签节点、视频偏好标签节点、游戏偏好标签节点中的一种或多种；每一子节点连接多个尾节点。

[0042] 以上任一实施例方案至少具有以下有益效果：

[0043] 应用本实施例方法，能够生成包括用户的基础属性、娱乐行为、驾驶行为、消费行

为等维度更为丰富的用户画像,可以让车厂更全面了解车辆用户的需求/习惯,优化下一代车型的设计,为开发符合车辆用户需求的应用服务和产品推荐提供有力支撑。

[0044] 本发明的其它特征和优点将在随后的说明书中阐述,并且,部分地从说明书中变得显而易见,或者通过实施本发明而得以体现。本发明的目的和其他优点可通过在说明书、权利要求书以及附图中所特别指出的结构来实现和获得。

## 附图说明

[0045] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0046] 图1为本发明一实施例所述一种车联网用户画像生成方法的流程示意图。

[0047] 图2为本发明一实施例中标签树模型的结构示意图。

[0048] 图3为本发明另一实施例中一种车联网用户画像生成系统的框架图。

## 具体实施方式

[0049] 以下将参考附图详细说明本公开的各种示例性实施例、特征和方面。附图中相同的附图标记表示功能相同或相似的元件。尽管在附图中示出了实施例的各种方面,但是除非特别指出,不必按比例绘制附图。

[0050] 另外,为了更好的说明本发明,在下文的具体实施例中给出了众多的具体细节。本领域技术人员应当理解,没有某些具体细节,本发明同样可以实施。在一些实例中,对于本领域技术人员熟知的手段未作详细描述,以便于凸显本发明的主旨。

[0051] 本发明一实施例提出一种车联网用户画像生成方法,图1为本实施例方法的流程图,参阅图1,本实施例方法包括以下步骤S1-S3:

[0052] 步骤S1、根据用户身份信息获取车联网中多个数据源的原始用户数据,每一数据源的原始用户数据包括用户的基础属性数据、消费行为数据、驾驶行为数据、娱乐行为数据中的一种或多种;

[0053] 具体而言,所述多个数据源包括但不限于为车辆Tbox端、安装于用户手机上的用于控制/管理车辆的APP端、车机AVNT端、B端业务系统(例如产线检测系统、新能源转发系统等)以及第三方生态应用端(例如用户所使用的腾讯QQ音乐、酷狗音乐、喜马拉雅、企鹅FM、车载微信等应用)的数据。可以理解的是,每一个数据源的原始用户数据都与用户身份信息相关联,用户身份信息例如是账号信息、手机号码等,根据目标用户的身份信息即可以获取对应的目标用户的车联网中多个数据源的原始用户数据。

[0054] 其中,所述基础属性数据包括用户的性别信息、年龄信息、地域信息、学历信息、收入信息中的一种或多种;

[0055] 其中,所述消费行为数据包括购物信息、停车信息、加油信息、充电信息、流量充值信息、积分兑换信息中的一种或多种;

[0056] 其中,所述驾驶行为数据包括油耗信息、速度信息、里程信息、电耗信息、驾驶行为评分信息、保险信息中的一种或多种;

[0057] 其中,所述娱乐行为数据包括音乐偏好信息、电台偏好信息、视频偏好信息、游戏偏好信息中的一种或多种。

[0058] 步骤S2、按照预设处理规则对所述多个数据源的原始用户数据进行预处理得到待标记用户数据;

[0059] 具体而言,上述多个数据源的原始用户数据包含了大量数据,那么这些数据有可能是存在错误的数据,因此,需要对原始用户数据进行预处理之后才能够用于后续的标签标记,预处理为数据清理,具体包括检查数据一致性,处理无效值和缺失值等,数据清洗具体可以利用有关技术,例如数理统计、数据挖掘或预定义的清理规则将脏数据转化为满足数据质量要求的数据,即待标记用户数据。

[0060] 步骤S3、按照预设标签规则对所述待标记用户数据进行标签标记得到用户标签数据,所述用户标签数据包括多个用户数据及其标签;

[0061] 具体而言,在应用本实施例方法之前,预先设定一个标签规则,即某一类型/属性的数据,对应关联某一种标签,例如,年龄段标签:25~35岁;又例如,地域标签:北京;因此,根据所述预设标签规则,可以为每一个用户数据标记对应的标签。

[0062] 步骤S4、对所述用户标签数据进行聚类分析得到多个用户数据集合;其中每一用户数据集合包括相似的多个用户数据及其标签;

[0063] 具体而言,聚类是将数据分类到不同的类或者簇这样的一个过程,所以同一个簇中的对象有很大的相似性,而不同簇间的对象有很大的相异性,所述簇即本实施例中的用户数据集合。因此,通过聚类分析,可以将相似的多个用户数据划分到相同的簇中。

[0064] 需说明的是,聚类的方式有很多种,本实施例中不限于某一种聚类方式。

[0065] 其中,步骤S4的目的在于对标记相应标签的大量用户数据进行统计,以便于提高后续利用用户数据标签对用户画像进行多维度描述的效率。

[0066] 步骤S5、根据所述多个用户数据集合计算每一标签的比重和稀缺度,并根据所述每一标签的比重和稀缺度对用户进行多维度描述,以生成用户画像。

[0067] 具体而言,用户画像又称用户角色,作为一种勾画目标用户、联系用户诉求与设计方向的有效工具,用户画像在各领域得到了广泛的应用。我们在实际操作的过程中往往将以最为浅显和贴近生活的话语将用户的属性、行为与期待的数据转化联结起来。作为实际用户的虚拟代表,用户画像所形成的用户角色并不是脱离产品和市场之外所构建出来的,形成的用户角色需要有代表性能代表产品的主要受众和目标群体。其中,根据所述多个用户数据集合,可以确定关于与目标用户有关的标签情况,例如某一标签所出现的频率,以及逆向频率(稀缺度),据此,可以根据标签所出现的频率,以及逆向频率对该标签给予适当的描绘,例如,标签尺寸大小,透明度,等等,描绘的方式可以自行设定。

[0068] 在一实施例中,所述步骤S4具体包括:

[0069] S41、随机确定一个k值;

[0070] S42、从所述多个用户数据中随机选择k个用户数据作为质心;

[0071] S43、对所述多个用户数据中的每一个用户数据,计算其与每一个质心的距离,并将其划分至与其距离最小的质心所属的用户数据集合,以得到k个用户数据集合;

[0072] 其中,所述距离为欧氏距离,或曼哈顿距离,或切比雪夫距离。

[0073] S44、重新计算所述k个用户数据集合的质心;

[0074] S45、如果重新计算出来的用户数据集合质心和原来的用户数据集合质心之间的距离小于预设阈值，表示重新计算的质心的位置变化不大，趋于稳定，或者说收敛，可以认为聚类已经达到期望的结果，则结束聚类；

[0075] 如果重新计算出来的用户数据集合质心和原来的用户数据集合质心之间的距离大于等于预设阈值，令 $k=k+1$ ，并返回步骤S42继续聚类，循环步骤S42-S45，直至如果重新计算出来的用户数据集合质心和原来的用户数据集合质心之间的距离小于预设阈值。

[0076] 在一实施例中，所述步骤S5包括：

[0077] 步骤S511、根据公式 $TF_i = W_i/N_1$ 计算每一用户标签的比重；其中， $TF_i$ 为用户标签*i*的比重， $W_i$ 为用户标签*i*所对应的用户数据集合中用户标签*i*的个数， $N_1$ 为多个用户数据集合的标签总数；

[0078] 步骤S512、根据公式 $IDF_i = M_i/N_2$ 计算每一用户标签的稀缺度；其中， $IDF_i$ 为用户标签*i*的稀缺度， $M_i$ 为用户数据集合的个数， $N_2$ 为包含用户标签*i*的用户数据集合的个数；

[0079] 步骤S513、根据每一用户标签的比重和稀缺度分别对每一用户标签进行多维度描述，并可视化描述内容，以生成用户画像。

[0080] 在一实施例中，所述方法还包括：根据所述用户标签数据构建标签树模型，在所述标签树模型中，若某一用户数据集合包括某一标签，则该用户数据集合作为一个尾节点被连接至对应的上层标签节点；

[0081] 在本实施例中，所述步骤S5具体包括：

[0082] 根据所述多个用户数据集合以及所述标签树模型计算每一标签的比重和稀缺度，并根据所述每一标签的比重和稀缺度对用户进行多维度描述，以生成用户画像。

[0083] 在一实施例中，如图2所示，所述标签树模型的首节点包括基础属性节点、消费行为节点、驾驶行为节点以及娱乐行为节点；

[0084] 其中，所述基础属性节点的子节点包括性别标签节点、年龄标签节点、地域标签节点、学历标签节点、收入标签节点中的一个或多个，所述消费行为节点的子节点包括购物标签节点、停车标签节点、加油标签节点、充电标签节点、流量充值标签节点、积分兑换标签节点中的一种或多种；所述驾驶行为节点的子节点包括油耗标签节点、速度标签节点、里程标签节点、电耗标签节点、驾驶行为评分标签节点、保险标签节点中的一种或多种；所述娱乐行为节点的子节点包括音乐偏好标签节点、电台偏好标签节点、视频偏好标签节点、游戏偏好标签节点中的一种或多种；每一子节点连接多个尾节点。

[0085] 在一实施例中，所述步骤S5具体包括：

[0086] 步骤S521、根据公式 $TF_i = W_i/N_1$ 计算每一用户标签的比重；其中， $TF_i$ 为用户标签*i*的比重， $W_i$ 为用户标签*i*所对应的用户数据集合中用户标签*i*的个数， $N_1$ 为多个用户数据集合的标签总数；

[0087] 步骤S522、根据公式 $IDF_i = M_i/N_2$ 计算每一用户标签的稀缺度；其中， $IDF_i$ 为用户标签*i*的稀缺度， $M_i$ 为用户数据集合的个数， $N_2$ 为包含用户标签*i*所对应的标签子节点所连接的尾节点个数；

[0088] 步骤S523、根据每一用户标签的比重和稀缺度分别对每一用户标签进行多维度描述，并可视化描述内容，以生成用户画像。

[0089] 通过以上实施例的描述可知，本实施例方法能够生成包括用户的基础属性、娱乐

行为、驾驶行为、消费行为等维度更为丰富的用户画像,可以让车厂更全面了解车辆用户的需求/习惯,优化下一代车型的设计,为开发符合车辆用户需求的应用服务和产品推荐提供有力支撑。

[0090] 参阅图3,本发明另一实施例还提出一种车联网用户画像生成系统,包括:

[0091] 原始数据获取单元1,用于根据用户身份信息获取车联网中多个数据源的原始用户数据,每一数据源的原始用户数据包括用户的基础属性数据、消费行为数据、驾驶行为数据、娱乐行为数据中的一种或多种;

[0092] 预处理单元2,用于按照预设处理规则对所述多个数据源的原始用户数据进行预处理得到待标记用户数据;

[0093] 数据标记单元3,用于按照预设标签规则对所述待标记用户数据进行标签标记得到用户标签数据,所述用户标签数据包括多个用户数据及其标签;

[0094] 聚类分析单元4,用于对所述用户标签数据进行聚类分析得到多个用户数据集合;其中每一用户数据集合包括相似的多个用户数据及其标签;以及

[0095] 画像生成单元5,用于根据所述多个用户数据集合计算每一标签的比重和稀缺度,并根据所述每一标签的比重和稀缺度对用户进行多维度描述,以生成用户画像。

[0096] 在一实施例中,所述基础属性数据包括用户的性别信息、年龄信息、地域信息、学历信息、收入信息中的一种或多种;所述消费行为数据包括购物信息、停车信息、加油信息、充电信息、流量充值信息、积分兑换信息中的一种或多种;所述驾驶行为数据包括油耗信息、速度信息、里程信息、电耗信息、驾驶行为评分信息、保险信息中的一种或多种;所述娱乐行为数据包括音乐偏好信息、电台偏好信息、视频偏好信息、游戏偏好信息中的一种或多种。

[0097] 在一实施例中,所述画像生成单元5包括:

[0098] 标签树构建单元51,用于根据所述用户标签数据构建标签树模型,在所述标签树模型中,若某一用户数据集合包括某一标签,则该用户数据集合作为一个尾节点被连接至对应的上层标签节点;以及

[0099] 画面描述单元52,用于根据所述多个用户数据集合以及所述标签树模型对用户进行多维度描述,以生成用户画像;

[0100] 其中,所述标签树模型的首节点包括基础属性节点、消费行为节点、驾驶行为节点以及娱乐行为节点;所述基础属性节点的子节点包括性别标签节点、年龄标签节点、地域标签节点、学历标签节点、收入标签节点中的一个或多个,所述消费行为节点的子节点包括购物标签节点、停车标签节点、加油标签节点、充电标签节点、流量充值标签节点、积分兑换标签节点中的一种或多种;所述驾驶行为节点的子节点包括油耗标签节点、速度标签节点、里程标签节点、电耗标签节点、驾驶行为评分标签节点、保险标签节点中的一种或多种;所述娱乐行为节点的子节点包括音乐偏好标签节点、电台偏好标签节点、视频偏好标签节点、游戏偏好标签节点中的一种或多种;每一子节点连接多个尾节点。

[0101] 以上所描述的系统实施例仅仅是示意性的,其中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模块来实现本实施例方案的目的。

[0102] 需说明的是,上述实施例所述系统与上述实施例所述方法对应,因此,上述实施例所述系统未详述部分可以参阅上述实施例所述方法的内容得到,此处不再赘述。

[0103] 并且,上述实施例所述车联网用户画像生成系统如果以软件功能单元的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储介质中。

[0104] 具体而言,所述计算机可读存储介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM, Read-Only Memory)、随机存取存储器(RAM, Random Access Memory)、电载波信号、电信信号以及软件分发介质等。

[0105] 以上已经描述了本发明的各实施例,上述说明是示例性的,并非穷尽性的,并且也不限于所披露的各实施例。在不偏离所说明的各实施例的范围和精神的情况下,对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择,旨在最好地解释各实施例的原理、实际应用或对市场中的技术改进,或者使本技术领域的其它普通技术人员能理解本文披露的各实施例。

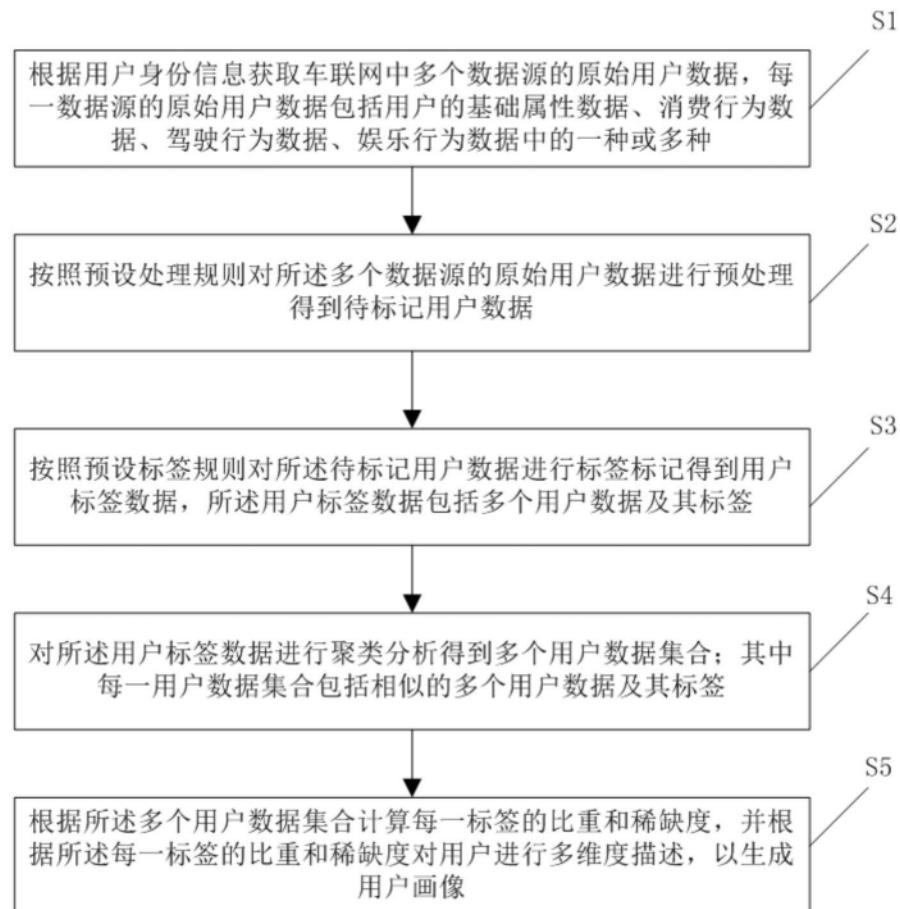


图1

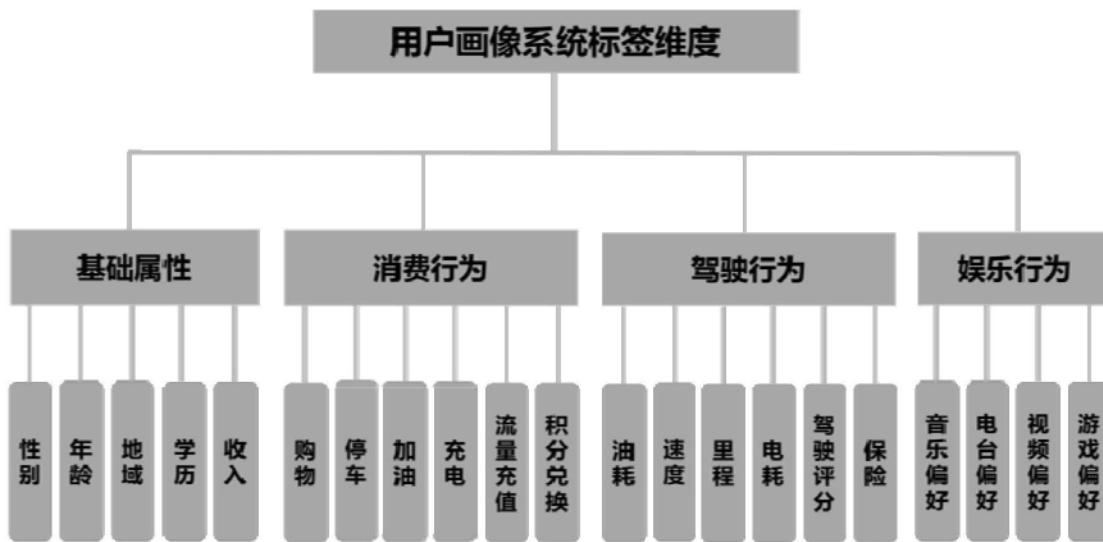


图2

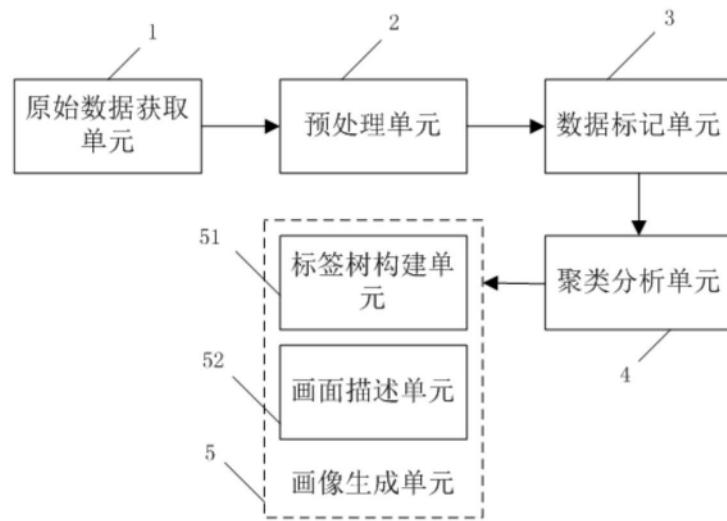


图3