



(12) 发明专利申请

(10) 申请公布号 CN 114511353 A

(43) 申请公布日 2022. 05. 17

(21) 申请号 202210106089.4

G06F 3/04847 (2022.01)

(22) 申请日 2022.01.28

(71) 申请人 中国工商银行股份有限公司

地址 100140 北京市西城区复兴门内大街
55号

(72) 发明人 陆怡 贾玉红 徐聿帆 林孙镇江

(74) 专利代理机构 北京同立钧成知识产权代理
有限公司 11205

专利代理师 余娜 臧建明

(51) Int. Cl.

G06Q 30/02 (2012.01)

G06F 16/54 (2019.01)

G06F 16/53 (2019.01)

G06F 16/36 (2019.01)

G06F 16/18 (2019.01)

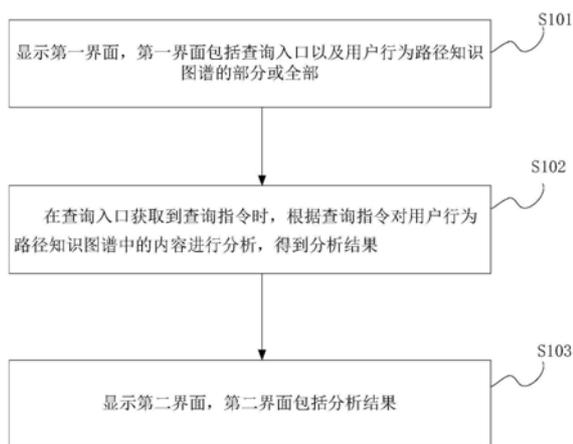
权利要求书4页 说明书16页 附图13页

(54) 发明名称

数据分析方法和装置

(57) 摘要

本申请实施例提供一种数据分析方法和装置,涉及大数据技术领域,包括:显示第一界面,第一界面包括查询入口以及用户行为路径知识图谱的部分或全部;在查询入口获取到查询指令时,根据查询指令对用户行为路径知识图谱中的内容进行分析,得到分析结果;显示第二界面,第二界面包括分析结果。本申请的数据分析方法采用知识图谱技术对用户行为路径进行分析,可以呈现出更加丰富的信息,对于直观了解整个用户行为路径、钻取挖掘用户行为特征、分析路径缺陷和提升产品价值,具有重要的意义。



1. 一种数据分析方法,其特征在于,包括:

显示第一界面,所述第一界面包括查询入口以及用户行为路径知识图谱的部分或全部;所述用户行为路径知识图谱包括多个节点、所述多个节点之间的关系、任一个所述节点的节点属性以及任一个所述关系的关系属性;所述节点属性包括节点类型或点击次数的一种或多种,所述关系属性包括用户行为路径的路径频次、路径类型或路径发生时间的一种或多种;

在所述查询入口获取到查询指令时,根据所述查询指令对所述用户行为路径知识图谱中的内容进行分析,得到分析结果;

显示第二界面,所述第二界面包括所述分析结果。

2. 根据权利要求1所述的方法,其特征在于,所述查询入口包括封装的可视化查询接口,和/或,用于写入查询语言的写入区域;

所述查询指令包括下述的一项或多项:最短路径查询指令、节点扩展指令、节点合并指令、节点拆分指令、按节点属性筛选条件过滤节点指令、关系合并指令、关系拆分指令、按关系属性筛选条件过滤关系指令、关系权重计算指令或图指标计算指令。

3. 根据权利要求2所述的方法,其特征在于,所述在所述查询入口获取到查询指令时,根据所述查询指令对所述用户行为路径知识图谱中的内容进行分析,得到分析结果,包括:

在所述查询入口获取到所述节点合并指令时,将所述节点合并指令对应的多个节点进行合并,以及将所述节点合并指令对应的多个节点之间的关系进行合并,得到所述分析结果,所述分析结果包括节点合并后的用户行为路径指示图谱;

或者,在所述查询入口获取到所述节点拆分指令时,将合并后的节点重新恢复为可查看明细节点信息的状态,以及将所述拆分指令对应的节点之间的关系进行拆分,得到所述分析结果,所述分析结果包括节点拆分后的用户行为路径指示图谱;

或者,在所述查询入口获取到所述按节点属性筛选条件过滤节点指令时,隐藏不符合所述节点属性筛选条件的节点和对应的关系,得到所述分析结果,所述分析结果包括隐藏部分节点的用户行为路径指示图谱;

或者,在所述查询入口获取到所述关系合并指令时,将所述关系合并指令对应的节点之间的多个关系进行合并,得到所述分析结果,所述分析结果包括关系合并后的用户行为路径指示图谱;

或者,在所述查询入口获取到所述关系拆分指令时,将合并后的关系重新恢复为可查看明细关系信息的状态,得到所述分析结果,所述分析结果包括关系拆分后的用户行为路径指示图谱;

或者,在所述查询入口获取到所述按关系属性筛选条件过滤关系指令时,隐藏不符合所述关系属性筛选条件的节点和对应的关系,得到所述分析结果,所述分析结果包括隐藏部分节点的用户行为路径指示图谱;

或者,在所述查询入口获取到所述关系权重计算指令时,计算某一路径内部数量占两点间全部路径数量的比重,并按照所述比重对所述关系进行粗细标记,所述比重越大,所述关系之间的线条越粗,得到所述分析结果,所述分析结果包括对关系标注后的用户行为路径指示图谱;

或者,在所述查询入口获取到所述图指标计算指令时,对所述用户行为路径知识图谱

进行计算,得到链路文件,所述分析结果包括所述链路文件。

4. 根据权利要求2或3所述的方法,其特征在于,所述查询指令还包括:节点统计指令和关系统计指令中的至少一项;所述在所述查询入口获取到查询指令时,根据所述查询指令对所述用户行为路径知识图谱中的内容进行分析,得到分析结果,包括:

在所述查询入口获取到所述节点统计指令时,根据节点标签或节点属性的名称统计所述用户行为路径知识图谱中各节点在所述用户行为路径知识图谱中的数量或本次查询出的数量,或者,根据节点标签或节点属性的值,统计所述用户行为路径知识图谱中各节点在所述用户行为路径知识图谱中的分布情况或本次查询结果中不同值的分布状况,或者,根据目标节点中某个具体的属性,统计所述目标节点在所述用户行为路径知识图谱中的值的分布情况或本次查询结果中值的分布情况,或者,根据节点标签所属的功能模块划分多个社群,统计不同社群的内部的节点分布状况和对应的数量信息;

或者,在所述查询入口获取到所述关系统计指令时,根据关系标签或关系属性的名称,统计所述用户行为路径知识图谱中各节点在所述用户行为路径知识图谱中的数量或本次查询出的数量,或者,根据关系标签或关系属性的值,统计所述用户行为路径知识图谱中各节点在所述用户行为路径知识图谱中的分布情况或本次查询结果中不同值的分布状况,或者,根据目标关系中某个具体的属性,统计所述目标关系在所述用户行为路径知识图谱中的值的分布情况或本次查询结果中值的分布情况,或者,根据 3σ 原则,统计主流路径和小众路径的集合和分布数量,按照全量显示的原则,突出显示全部的非主流路径,或者,统计一定条件下所有的逆向路径;其中,所述主流路径为路径通过频率大于或等于频率阈值的路径,所述小众路径为路径通过频率小于所述频率阈值的路径,所述非法路径为超出程序允许范围的用户操作路径。

5. 根据权利要求1-3任一项所述的方法,其特征在于,所述第一界面还包括滑动时间窗,所述方法还包括:

当接收到对所述滑动时间窗的起止时间修改时,所述第一界面显示修改后的所述滑动时间窗的时间内的部分所述用户行为路径知识图谱。

6. 根据权利要求1-3任一项所述的方法,其特征在于,所述显示第一界面包括:

向图数据库发送post请求;

获取所述post请求的结果;所述结果为js对象简谱json字符串数据;

将所述结果转化为可视化所需的json数据格式;

基于预设的可视化渲染的技术框架、力导向布局、可视化样式配置信息以及对所述可视化所需的json数据格式的结果进行可视化渲染,得到可视化结果;其中,所述可视化结果支持节点拖拽、节点关系的属性交互显示、右击节点后扩展的一种或多种功能。

7. 根据权利要求6所述的方法,其特征在于,所述可视化样式配置信息包括:节点样式配置信息和关系样式配置信息;

所述节点样式配置信息包括:节点文字显示内容、节点大小、节点颜色、节点图标的一种或多种;

所述关系样式配置信息包括:关系文字显示内容、关系线条粗细、关系线条颜色、关系线条线型的一种或多种。

8. 根据权利要求1-3任一项所述的方法,其特征在于,还包括:

收集用户行为日志；

将所述用户行为日志分类进行实体数据加工和关系数据加工，得到实体数据和关系数据；

基于所述实体数据和所述关系数据，构建所述用户行为路径知识图谱。

9. 根据权利要求8所述的方法，其特征在于，所述用户行为日志包括埋点日志和路由日志；所述实体数据加工包括：实体标签加工和实体属性加工；所述关系数据加工包括：关系标签加工和关系属性加工；

其中，所述实体标签加工包括：埋点实体标签加工和路由实体标签加工；所述埋点实体标签加工中将埋点实体标签按照不同的实体业务需求划分得到不同的实体，所述按照不同的实体业务需求划分包括下述的一种或多种：按照控件类型划分、按照不同点击对象的功能划分、按照业务功能划分等；所述路由实体标签加工中将路由实体标签按照下述的一种或多种进行划分：按照页面深度划分、按照页面广度划分、按照页面功能划分；

所述实体属性加工包括：将经过所述实体标签加工后得到的标签按照不同的业务需求划分为多种第一标签，将所述多种第一标签中的一种作为所述实体标签，以及将所述多种第一标签中除所述实体标签外的其余信息作为所述实体属性信息；

所述关系标签加工中按照不同的关系业务需求划分得到不同的实体；所述按照不同的关系业务需求划分包括下述的一种或多种：按照路径频率划分、按照路径价值划分、按照路径驻留时间划分、按照路径起始深度是否相同划分、路径起始深度相同时，按照是否由浅向深访问划分；

所述关系属性加工包括：将经过所述关系标签加工后得到的标签按照不同的业务需求划分为多种第二标签，将所述多种第二标签中的一种作为所述关系标签，以及将所述多种第二标签中除所述关系标签外的其余信息作为所述关系属性信息。

10. 根据权利要求9所述的方法，其特征在于，基于所述实体数据和所述关系数据，构建所述用户行为路径知识图谱，包括：

利用所述实体数据和所述关系数据进行本体构建和知识映射，得到所述用户行为路径知识图谱；

其中，所述本体构建包括：利用所述实体标签构建节点，所述实体标签与所述节点一一对应，各所述节点对应有各自的实体属性信息；利用所述关系标签构建关系，所述关系标签与所述关系一一对应，各所述关系对应有各自的关系属性信息；

所述知识映射包括：根据所述本体构建中的所述实体标签、所述关系标签及所述实体标签和所述关系标签各自对应的所述属性信息，与实际数据处理中数据文件的字段进行一一映射。

11. 一种数据分析装置，其特征在于，包括：

第一显示模块，用于显示第一界面，所述第一界面包括查询入口以及用户行为路径知识图谱的部分或全部；所述用户行为路径知识图谱包括多个节点、所述多个节点之间的关系、任一个所述节点的节点属性以及任一个所述关系的关系属性；所述节点属性包括节点类型、点击次数的一种或多种，所述关系属性包括用户行为路径的路径频次、路径类型或路径发生时间的一种或多种；

分析模块，用于在所述查询入口接收到查询指令时，根据所述查询指令对所述用户行

为路径知识图谱中的内容进行分析,得到分析结果;

第二显示模块,用于显示第二界面,所述第二界面包括所述分析结果。

12. 一种电子设备,其特征在于,包括:

存储器,用于存储计算机程序;

处理器,用于执行所述计算机程序以实现权利要求1至10中任一项所述方法。

13. 一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至10中任一项所述的方法的步骤。

数据分析方法和装置

技术领域

[0001] 本申请涉及大数据技术领域,特别涉及一种数据分析方法和装置。

背景技术

[0002] 在大数据时代,收集用户数据并对其进行分析可以帮助开发商改善与优化产品,为用户提供更加完善的产品服务。

[0003] 用户在操作网页或应用程序(application,app)的过程中,终端设备可以记录用户的每一次操作行为,并生成对应的日志。产品经理或运营分析人员可以基于用户行为生成的日志,对用户行为进行分析。例如,对用户行为进行分析时,可以分析得到用户行为路径,以通过用户行为路径实现刻画用户行为特征或者优化和改进产品设计等。然而,现有的用户行为路径分析方法存在分析结果不全面及不准确的问题。

发明内容

[0004] 本申请实施例提供一种数据分析方法和装置,可以全面的收集用户行为路径数据,提升用户行为路径的分析能力及分析结果的准确性。

[0005] 第一方面,本申请实施例提供一种数据分析方法,包括:

[0006] 显示第一界面,第一界面包括查询入口以及用户行为路径知识图谱的部分或全部;用户行为路径知识图谱包括多个节点、多个节点之间的关系、任一个节点的节点属性以及任一个关系的关系属性;节点属性包括节点类型或点击次数的一种或多种,关系属性包括用户行为路径的路径频次、路径类型或路径发生时间的一种或多种;

[0007] 在查询入口获取到查询指令时,根据查询指令对用户行为路径知识图谱中的内容进行分析,得到分析结果;

[0008] 显示第二界面,第二界面包括分析结果。

[0009] 可选的,查询入口包括封装的可视化查询接口,和/或,用于写入查询语言的写入区域。

[0010] 查询指令包括下述的一项或多项:最短路径查询指令、节点扩展指令、节点合并指令、节点拆分指令、按节点属性筛选条件过滤节点指令、关系合并指令、关系拆分指令、按关系属性筛选条件过滤关系指令、关系权重计算指令或图指标计算指令。

[0011] 可选的,在查询入口获取到查询指令时,根据查询指令对用户行为路径知识图谱中的内容进行分析,得到分析结果,包括:

[0012] 在查询入口获取到节点合并指令时,将节点合并指令对应的多个节点进行合并,以及将节点合并指令对应的多个节点之间的关系进行合并,得到分析结果,分析结果包括节点合并后的用户行为路径指示图谱;

[0013] 或者,在查询入口获取到节点拆分指令时,将合并后的节点重新恢复为可查看详细节点信息的状态,以及将拆分指令对应的节点之间的关系进行拆分,得到分析结果,分析结果包括节点拆分后的用户行为路径指示图谱;

[0014] 或者,在查询入口获取到按节点属性筛选条件过滤节点指令时,隐藏不符合节点属性筛选条件的节点和对应的关系,得到分析结果,分析结果包括隐藏部分节点的用户行为路径指示图谱;

[0015] 或者,在查询入口获取到关系合并指令时,将关系合并指令对应的节点之间的多个关系进行合并,得到分析结果,分析结果包括关系合并后的用户行为路径指示图谱;

[0016] 或者,在查询入口获取到关系拆分指令时,将合并后的关系重新恢复为可查看详细关系信息的状态,得到分析结果,分析结果包括关系拆分后的用户行为路径指示图谱;

[0017] 或者,在查询入口获取到按关系属性筛选条件过滤关系指令时,隐藏不符合关系属性筛选条件的节点和对应的关系,得到分析结果,分析结果包括隐藏部分节点的用户行为路径指示图谱;

[0018] 或者,在查询入口获取到关系权重计算指令时,计算某一路径内部数量占两点间全部路径数量的比重,并按照比重对关系进行粗细标记,比重越大,关系之间的线条越粗,得到分析结果,分析结果包括对关系标注后的用户行为路径指示图谱;

[0019] 或者,在查询入口获取到图指标计算指令时,对用户行为路径知识图谱进行计算,得到链路文件,分析结果包括所述链路文件。

[0020] 可选的,查询指令还包括:节点统计指令和关系统计指令中的至少一项;在查询入口获取到查询指令时,根据查询指令对所述用户行为路径知识图谱中的内容进行分析,得到分析结果,包括:

[0021] 在查询入口获取到节点统计指令时,根据节点标签或节点属性的名称统计所述用户行为路径知识图谱中各节点在用户行为路径知识图谱中的数量或本次查询出的数量,或者,根据节点标签或节点属性的值,统计用户行为路径知识图谱中各节点在所述用户行为路径知识图谱中的分布情况或本次查询结果中不同值的分布状况,或者,根据目标节点中某个具体的属性,统计目标节点在用户行为路径知识图谱中的值的分布情况或本次查询结果中值的分布情况,或者,根据节点标签所属的功能模块划分多个社群,统计不同社群的内部的节点分布状况和对应的数量信息。

[0022] 或者,在查询入口获取到关系统计指令时,根据关系标签或关系属性的名称,统计用户行为路径知识图谱中各节点在用户行为路径知识图谱中的数量或本次查询出的数量,或者,根据关系标签或关系属性的值,统计用户行为路径知识图谱中各节点在用户行为路径知识图谱中的分布情况或本次查询结果中不同值的分布状况,或者,根据目标关系中某个具体的属性,统计目标关系在用户行为路径知识图谱中的值的分布情况或本次查询结果中值的分布情况,或者,根据 3σ 原则,统计主流路径和小众路径的集合和分布数量,按照全量显示的原则,突出显示全部的非主流路径,或者,统计一定条件下所有的逆向路径;其中,主流路径为路径通过频率大于或等于频率阈值的路径,小众路径为路径通过频率小于所述频率阈值的路径,非法路径为超出程序允许范围的用户操作路径。

[0023] 可选的,第一界面还包括滑动时间窗,数据分析方法还包括:

[0024] 当接收到对滑动时间窗的起止时间修改时,第一界面显示修改后的滑动时间窗的时间内的部分用户行为路径知识图谱。

[0025] 可选的,显示第一界面包括:

[0026] 向图数据库发送post请求;

- [0027] 获取所述post请求的结果;请求结果为js对象简谱json字符串数据;
- [0028] 将请求结果转化为可视化所需的json数据格式;
- [0029] 基于预设的可视化渲染的技术框架、力导向布局、可视化样式配置信息以及对可视化所需的json数据格式的结果进行可视化渲染,得到可视化结果;其中,可视化结果支持节点拖拽、节点关系的属性交互显示、右击节点后扩展的一种或多种功能。
- [0030] 可选的,本申请提供的数据分析方法还包括:
- [0031] 收集用户行为日志;
- [0032] 将用户行为日志分类进行实体数据加工和关系数据加工,得到实体数据和关系数据;
- [0033] 基于实体数据和关系数据,构建用户行为路径知识图谱。
- [0034] 可选的,用户行为日志包括埋点日志和路由日志;实体数据加工包括:实体标签加工和实体属性加工;关系数据加工包括:关系标签加工和关系属性加工;
- [0035] 其中,实体标签加工包括:埋点实体标签加工和路由实体标签加工;埋点实体标签加工中将埋点实体标签按照不同的实体业务需求划分得到不同的实体,按照不同的实体业务需求划分包括下述的一种或多种:按照控件类型划分、按照不同点击对象的功能划分、按照业务功能划分等;所述路由实体标签加工中将路由实体标签按照下述的一种或多种进行划分:按照页面深度划分、按照页面广度划分、按照页面功能划分;
- [0036] 实体属性加工包括:将经过实体标签加工后得到的标签按照不同的业务需求划分为多种第一标签,将多种第一标签中的一种作为实体标签,以及将多种第一标签中除所述实体标签外的其余信息作为所述实体属性信息;
- [0037] 关系标签加工中按照不同的关系业务需求划分得到不同的实体;按照不同的关系业务需求划分包括下述的一种或多种:按照路径频率划分、按照路径价值划分、按照路径驻留时间划分、按照路径起始深度是否相同划分、路径起始深度相同时,按照是否由浅向深访问划分;
- [0038] 关系属性加工包括:将经过关系标签加工后得到的标签按照不同的业务需求划分为多种第二标签,将多种第二标签中的一种作为所述关系标签,以及将多种第二标签中除关系标签外的其余信息作为所述关系属性信息。
- [0039] 可选的,基于实体数据和关系数据,构建所述用户行为路径知识图谱,包括:
- [0040] 利用实体数据和关系数据进行本体构建和知识映射,得到用户行为路径知识图谱;
- [0041] 其中,本体构建包括:利用实体标签构建节点,实体标签与节点一一对应,各节点对应有各自的实体属性信息;利用关系标签构建关系,关系标签与关系一一对应,各关系对应有各自的关系属性信息;
- [0042] 知识映射包括:根据本体构建中的实体标签、关系标签及实体标签和关系标签各自对应的所述属性信息,与实际数据处理中数据文件的字段进行一一映射。
- [0043] 第二方面,本申请提供一种数据分析装置,包括:
- [0044] 第一显示模块,用于显示第一界面,第一界面包括查询入口以及用户行为路径知识图谱的部分或全部;用户行为路径知识图谱包括多个节点、多个节点之间的关系、任一个节点的节点属性以及任一个关系的关系属性;节点属性包括节点类型、点击次数的一种或

多种,关系属性包括用户行为路径的路径频次、路径类型或路径发生时间的一种或多种;

[0045] 分析模块,用于在查询入口接收到查询指令时,根据查询指令对用户行为路径知识图谱中的内容进行分析,得到分析结果;

[0046] 第二显示模块,用于显示第二界面,第二界面包括分析结果。

[0047] 第三方面,本申请实施例提供一种电子设备,包括:存储器和处理器;

[0048] 存储器用于存储计算机程序;处理器用于运行存储器存储的计算机程序实现第一方面中任一项的方法。

[0049] 第四方面,本申请实施例提供一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行以实现第一方面中任一项的方法。

[0050] 第五方面,本申请实施例提供一种计算机程序产品,包括计算机程序,该计算机程序被处理器执行时实现第一方面中任一项的方法。

[0051] 本申请提供的数据分析方法及装置,在数据分析系统接收到用户输入的查询指令后,可以对用户的行为路径知识图谱中的内容进行分析,获得分析结果。由于采用知识图谱技术对用户行为路径进行分析,相比于现有的用户行为路径分析方法,不仅可以对用户行为路径进行单向分析,也可以对用户行为路径中的反向路径和页面间反复横跳的用户行为路径进行分析,获得更加全面和更加准确的分析结果。这对于直观了解整个用户行为路径、钻取挖掘用户行为特征、分析路径缺陷和提升产品价值,具有重要的意义。

附图说明

[0052] 图1为本申请实施例提供的场景示意图;

[0053] 图2为本申请实施例的提供的数据分析方法的流程示意图;

[0054] 图3为本申请实施例的提供的第一界面的示意图;

[0055] 图4为本申请实施例的提供的用户触发查询控件的界面示意图;

[0056] 图5为本申请实施例的提供的用户触发节点合并控件的界面示意图;

[0057] 图6为本申请实施例的提供的用户触发节点拆分控件的界面示意图;

[0058] 图7为本申请实施例的提供的用户触发按条件过滤节点控件的界面示意图;

[0059] 图8为本申请实施例的提供的用户触发关系合并控件的界面示意图;

[0060] 图9为本申请实施例的提供的用户触发关系拆分控件的界面示意图;

[0061] 图10为本申请实施例的提供的用户触发关系权重计算控件的界面示意图;

[0062] 图11为本申请实施例的提供的用户触发节点统计控件的界面示意图;

[0063] 图12为本申请实施例的提供的第一界面的示意图;

[0064] 图13为本申请实施例的提供的可视化配置界面示意图;

[0065] 图14为本申请实施例的提供的数据分析方法的流程示意图;

[0066] 图15为本申请实施例的提供的数据分析系统的构成示意图;

[0067] 图16为本申请实施例的提供的数据分析装置示意图。

具体实施方式

[0068] 为了便于清楚描述本申请实施例的技术方案,以下,对本申请实施例中所涉及的部分术语和技术进行简单介绍:

[0069] 1) 埋点日志:埋点,就是在业务系统的程序中,植入一些收集事件数据的工具代码,进行各种事件的收集。埋点日志记录着特定流程的信息,用于跟踪功能的使用情况,是后续用来持续优化产品或提供运营的数据支撑。埋点日志是一种控件级别的日志,提供的信息主要包括访问数、访客数、停留时长、页面浏览数和跳出率等。

[0070] 2) 路由日志:记录路由器的工作状态的日志,记录着特定流程的页面跳转信息,用于跟踪页面整体的切换情况。路由日志是一种页面级别的日志,提供的信息主要包括页面响应时长、访问来源信息等。路由日志可提供宏观页面层面的用户行为路径数据。

[0071] 3) 知识图谱:是显示知识发展进程与结构关系的一系列各种不同的图形,用可视化技术描述知识资源及其载体,挖掘、分析、构建、绘制和显示知识及它们之间的相互联系。是通过将应用数学、图形学、信息可视化技术、信息科学等学科的理论与方法与计量学引文分析、共现分析等方法结合,并利用可视化的图谱形象地展示学科的核心结构、发展历史、前沿领域以及整体知识架构达到多学科融合目的的现代理论。

[0072] 4) 可视化:是利用计算机图形学和图像处理技术,将数据转换成图形或图像在屏幕上显示出来,再进行交互处理的理论、方法和技术。

[0073] 5) 用户行为路径:是指用户在操作网页或app的过程中形成的访问路径,例如,用户行为路径包括前一个埋点日志指向后一个埋点日志所形成的路径,或者前一个路由日志指向后一个路由日志所形成的路径。

[0074] 6) POST请求是向指定的资源提交要被处理的数据。

[0075] 7) 其他术语

[0076] 在本申请的实施例中,采用了“第一”、“第二”等字样对功能和作用基本相同的相同项或相似项进行区分。例如,第一界面、第二界面,并不对其先后顺序进行限定。本领域技术人员可以理解“第一”、“第二”等字样并不对数量和执行次序进行限定,并且“第一”、“第二”等字样也并不限定一定不同。

[0077] 需要说明的是,本申请实施例中,“示例性的”或者“例如”等词用于表示例子、例证或说明。本申请中被描述为“示例性的”或者“例如”的任何实施例或设计方案不应被解释为比其他实施例或设计方案更优选或更具优势。确切而言,使用“示例性的”或者“例如”等词旨在以具体方式呈现相关概念。

[0078] 本申请实施例中,“至少一个”是指一个或者多个,“多个”是指两个或两个以上。“和/或”,描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B的情况,其中A,B可以是单数或者复数。字符“/”一般表示前后关联对象是一种“或”的关系。“以下至少一项(个)”或其类似表达,是指的这些项中的任意组合,包括单项(个)或复数项(个)的任意组合。例如,a,b,或c中的至少一项(个),可以表示:a,b,c,a-b,a--c,b-c,或a-b-c,其中a,b,c可以是单个,也可以是多个。

[0079] 下面结合附图对本申请实施例提供的数据分析方法进行详细地介绍。需要说明的是,本申请实施例中的“在……时”,可以为在某种情况发生的瞬时,也可以为在某种情况发生后的一段时间内,本申请实施例对此不作具体限定。

[0080] 收集用户数据并对其进行分析可以帮助开发商改善与优化产品,为用户提供更加完善的产品服务。分析用户数据往往是对用户行为路径的分析。

[0081] 用户行为路径分析主要为,根据每位用户在操作过程中的点击行为日志,分析用

户在操作各个功能时的流转规律与特点,挖掘用户的访问或点击模式,进而解决一些特定的业务问题。例如提升核心功能的到达率、提取特定用户群体的主流路径行为、刻画用户浏览特征、优化和改进产品设计等。

[0082] 在不同类型的产品中,产品经理或运营分析人员基本都在使用漏斗图和桑基图作用户行为分析。其中,漏斗图的特点是目标明确、直观单一,适用链路单一的产品,可以对局部的转化问题做重点突破;而桑基图的特点是可以做时序分析且路径多元化,适用于分析主流路径和小众路径,常用场景为价值归因和个性化推荐。

[0083] 无论是漏斗图还是桑基图,都存在一种缺点:由于两种图在本质上都是单向图,因此较难观察和分析反向路径,也很难追踪页面间反复横跳的用户行为。而反向路径和反复横跳的用户行为往往代表着一些产品设计缺陷和有潜在价值的用户思维模式,观察分析这些路径和行为,对于优化产品设计、深入理解用户行为背后的心理因素和发掘用户对产品的价值期待,有着重要的意义。

[0084] 反向路径和反复横跳的用户行为本质上都是一种关联关系,而知识图谱是把不同种类的信息、知识连接在一起而得到的一个关系网络,提供了从“关系”的角度去分析问题的能力。进一步来看,知识图谱主要包括知识提取、知识表示、知识存储、知识挖掘、知识推理等相关技术,可以在关系发掘、群体识别、网络分析、事件传导等业务场景中使用。这些技术和适用场景完全满足了用户行为路径分析的要求。而且,知识图谱的可视化可以呈现出丰富的信息,这对于直观了解整个用户行为路径、钻取挖掘用户行为特征、分析路径缺陷和提升产品价值,具有重要的意义。

[0085] 有鉴于此,本申请提供一个用户行为路径覆盖全面(包含顺序路径和反向路径)、且支持路径灵活探查的用户行为路径的数据分析方法和装置,可以使数据分析人员直观了解整个用户行为路径,更高效便捷的运用大数据完成实际业务需求,提高数据挖掘的效率,进一步提升数据价值。

[0086] 下面以具体地实施例对本发明的技术方案以及本申请的技术方案如何解决上述技术问题进行详细说明。下面这几个具体的实施例可以相互结合,对于相同或相似的概念或过程可能在某些实施例中不再赘述。下面将结合附图,对本发明的实施例提供的数据分析方法进行描述。

[0087] 图1是本申请实施例提供场景图,如图1所示,包括用户A1和终端设备B1,终端设备可以接收用户对其的操作。终端设备搭载有对用户行为路径进行分析的数据分析系统,当终端设备接收到用户对数据分析系统的触发时,可以进入该数据分析系统,显示用于数据分析的界面。终端可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备,本申请实施例对此不做限定;终端设备可以接收用户对其的操作可以但不限于点击、触摸等,本申请实施例对此不做限定。

[0088] 在一个实施例中,如图2所示,提供了一种数据分析方法,以该方法应用于图1中的数据分析系统为例进行说明,包括以下步骤:

[0089] 步骤101:显示第一界面,第一界面包括查询入口以及用户行为路径知识图谱的部分或全部。

[0090] 其中,用户行为路径知识图谱包括多个节点、多个节点之间的关系、任一节点的节点属性以及任一关系的关系属性;节点属性包括节点类型、点击次数中的一种或多种,

关系属性包括用户行为路径的路径频次、路径类型或路径发生时间的一种或多种。

[0091] 本申请实施例中,用户行为路径知识图谱可以为使用知识图谱技术将用户在操作网页或app的过程中形成的访问路径串联起来,而得到的一个用户行为关系网络。例如,用户在页面A初停留的N分钟,从页面A操作转到页面B,从页面B返回到页面A又进入至页面C,利用知识图谱技术将上述的用户行为按时间先后顺序排列串联起来,得到的用户行为关系网络。

[0092] 其中,节点类型指埋点或者路由的类型;点击次数指一段时间内网站或者网页被访问的次数,通常以天为单位;用户行为路径的路径频次指用户通过该路径访问网站或页面的频率;路径类型指将用户访问网站或页面的路径划分为不同类型,例如,长时间驻留路径、短时间驻留路径等;路径发生时间指用户经过该路径访问网站或页面的时间。

[0093] 终端设备在接收到用户登录数据分析系统时,通过数据分析系统显示步骤101中的第一界面。

[0094] 第一界面如图3所示,其中,查询入口可以包括查询控件、统计控件和分析控件。数据分析系统可以接收到用户对其中任一控件的触发,数据分析系统在接收到用户对其中任一控件被触发时,生成对应的查询指令,数据分析系统根据查询指令对用户行为路径知识图谱中的内容进行分析。

[0095] 其中,数据分析系统在接收到用户对查询控件被触发时,可以对用户的行为路径知识图谱中的内容进行查询,获得查询结果;数据分析系统在接收到用户对统计控件被触发时,可以对用户的行为路径知识图谱中的内容进行统计,获得统计结果;数据分析系统在接收到用户对分析控件被触发时,可以对用户的行为路径知识图谱中的内容进行分析,获得分析结果。

[0096] 其中,第一界面还显示用户行为路径知识图谱的部分或全部,具体为,根据产品的不同,对应的用户行为路径知识图谱中的内容多少也存在不同,在接收到用户登录数据分析系统时,如果显示用户行为路径知识图谱的全部需要加载的时间较长,此时,显示用户行为路径知识图谱的部分;在接收到用户登录数据分析系统时,显示用户行为路径知识图谱的全部,需要的时间较短时,显示用户行为路径知识图谱的全部。显示用户行为路径知识图谱内容的数量,根据数据分析系统搭载的终端设备的不同而存在不同。

[0097] 步骤102:在查询入口获取到查询指令时,根据查询指令对用户行为路径知识图谱中的内容进行分析,得到分析结果。

[0098] 步骤103:显示第二界面,第二界面包括分析结果。

[0099] 数据分析系统在接收到步骤S102中的任一种查询指令后,数据分析系统在显示的第二界面中显示该指令对应的分析结果。

[0100] 其中,分析结果根据查询指令的不同显示不同的用户行为路径知识图谱,包括但不限于对用户行为路径知识图谱中内容的合并、拆分、隐藏或者标记。

[0101] 分析结果还可以以图表的形式展现给用户,包括但不限于柱状图、折线图、雷达图等,本申请不对图表的形式进行限定。可以理解的是,分析结果也可以使用其他方式进行展示,本申请也不对此进行限定。

[0102] 上述数据分析方法中,数据分析系统接收到用户输入的查询指令后,可以对用户的行为路径知识图谱中的内容进行分析,获得分析结果。由于采用知识图谱技术对用户行

为路径进行分析,相比于现有的用户行为路径分析方法,不仅可以对用户行为路径进行单向分析,也可以对用户行为路径中的反向路径和页面间反复横跳的用户行为路径进行分析,从而获得更加全面和更加准确的分析结果。这对于直观了解整个用户行为路径、钻取挖掘用户行为特征、分析路径缺陷和提升产品价值,具有重要的意义。

[0103] 可选的,在图2对应的实施例的基础上,下面对用户行为路径知识图谱中的内容进行分析的过程进行具体说明。

[0104] 可选的,查询入口包括封装的可视化查询接口,和/或,用于写入查询语言的写入区域。查询指令包括下述的一项或多项:最短路径查询指令、节点扩展指令、节点合并指令、节点拆分指令、按节点属性筛选条件过滤节点指令、关系合并指令、关系拆分指令、按关系属性筛选条件过滤关系指令、关系权重计算指令或图指标计算指令。

[0105] 数据分析系统接收到用户触发图3所示界面中的查询控件时,可显示封装的可视化查询接口,或者用于写入查询语言的写入区域,或者同时显示上述可视化查询接口和写入查询语言的写入区域。

[0106] 示例性的,如图4的a所示,数据分析系统接收到用户触发查询控件时,同时显示可视化查询接口和写入查询语言的写入区域。数据分析系统可以接收用户在查询语言的写入区域4010写入相应的查询语言,对用户行为路径知识图谱中的内容进行分析,获得分析结果;或者数据分析系统也可以接收用户在4011区域点击数据分析系统提供的一些将常用的查询逻辑固化后封装的可视化查询接口,对用户行为路径知识图谱中的内容进行分析,获得分析结果。

[0107] 可选的,在数据分析系统获得上述两种查询方式对应的分析结果后,数据分析系统支持对上述分析结果进行进一步查询,该查询方式称为扩展查询。扩展查询只能使用封装好的K-out查询接口或者固定的查询语句进行查询。此时,数据分析系统在接收到用户对查询控件的二次触发时,显示多个固定的查询接口供用户选择。例如,数据分析系统的查询结果显示节点A相关的信息后,以节点A为中心进行扩展查询,此时只能使用数据分析系统提供的查询接口进行查询,如点击节点A到节点F的最长路径的接口进行查询。

[0108] 其中,最短路径查询指令为查询两个节点之间的最短路径;节点扩展指令指获得某个节点的分析结果后,在此基础上进一步对该节点进行分析;节点合并指令为将多个节点合并为一个节点;节点拆分指令为将合并后的节点恢复原来的状态;按节点属性筛选条件过滤节点指令为按照一定的筛选条件,隐藏不符合条件的节点;关系合并指令将两个节点之间的关系进行合并;关系拆分指令将合并后的关系恢复原来的状态;按关系属性筛选条件过滤关系指令按照一定的筛选条件,隐藏不符合条件的关系;关系权重计算指令对两个节点之间的路径进行权重分析;图指标计算指令按照某一固定算法对用户行为路径知识图谱进行整体分析。

[0109] 可选的,在查询入口获取到查询指令时,根据查询指令对用户行为路径知识图谱中的内容进行分析,得到分析结果可以包括下述实现:

[0110] 可能的实现方式一:在查询入口获取到节点合并指令时,将节点合并指令对应的多个节点进行合并,以及将节点合并指令对应的多个节点之间的关系进行合并,得到分析结果,分析结果包括节点合并后的用户行为路径指示图谱。

[0111] 示例性的,查询指令可以包括有对应的指令控件,当数据分析系统接收到用户对

指令控件的触发时,生成对应查询指令,数据分析系统执行查询指令相对应的分析过程。

[0112] 示例性的,查询入口获取节点合并指令方法如下所示:

[0113] 如图5中a所示,当数据分析系统接收到用户对分析控件的触发时,显示二级界面,该二级界面包括节点分析控件、关系分析控件和图指标计算控件。当数据分析系统接收到用户对节点分析控件的触发时,显示如图5中b所示界面,该页面包括节点合并控件、节点拆分控件和按条件过滤节点控件。

[0114] 当数据分析系统接收到用户对节点合并控件的触发时,生成节点合并指令。其余查询指令的生成方式都与节点合并指令的生成方式相同或相似,后续不在赘述,都以用户触发相关指令控件来表示。

[0115] 如图5中c所示,当数据分析系统接收到用户对节点合并控件的触发时,可以显示下一级页面,该页面包括多个节点名称和一个合并控件,当数据分析系统接收到用户选中多个节点名称,并触发合并控件时,将选中的多个节点合并为一个节点。同时数据分析系统会自动将对应的节点之间的关系合并,生成合并后的用户行为路径指示图谱。

[0116] 可能的实现方式二,在查询入口获取到节点拆分指令时,将合并后的节点重新恢复为可查看明细节点信息的状态,以及将拆分指令对应的节点之间的关系进行拆分,得到分析结果,分析结果包括节点拆分后的用户行为路径知识图谱。

[0117] 如图6所示,数据分析系统接收到用户对节点拆分控件的触发时,显示下一级页面,该下一级页面中包括多个合并后的节点名称和一个拆分控件,当数据分析系统接收到用户选中某一合并后的节点名称,并触发拆分控件时,将合并后的节点进行拆分。节点拆分指令将合并后的节点重新恢复为可查看明细节点信息的状态的操作。拆分节点后,数据分析系统会将节点对应的关系也自动拆分,形成拆分后的用户行为路径知识图谱。

[0118] 可能的实现方式三,在查询入口获取到按节点属性筛选条件过滤节点指令时,隐藏不符合所述节点属性筛选条件的节点和对应的关系,得到分析结果,分析结果包括隐藏部分节点的用户行为路径指示图谱。

[0119] 如图7所示,数据分析系统接收到用户对按条件过滤节点控件的触发时,可以显示下一级页面,该下一级页面中包括可输入节点属性筛选条件的输入框和一个过滤控件。

[0120] 数据分析系统接收到用户在输入框中输入节点属性筛选条件,并触发过滤控件后,显示与输入的筛选条件有关的节点和节点之间的关系。按条件过滤节点支持层层过滤,即,在前一次过滤的基础上再次过滤。

[0121] 例如,节点A与节点C是与控件类型相关的节点,节点B是与控件类型无关的节点,数据分析系统在接收到用户在输入框中选择了控件类型并触发过滤指令后,对应的用户行为路径知识图谱中只显示节点A和节点C以及节点A与节点C之间的关系,节点A与节点B之间的关系、节点C与节点B之间的关系均被隐藏。

[0122] 可能的实现方式四,在查询入口获取到关系合并指令时,将关系合并指令对应的节点之间的多个关系进行合并,得到所述分析结果,所述分析结果包括关系合并后的用户行为路径指示图谱。

[0123] 关系合并往往被用于两节点之间关系数量较多的情况,如图8所示,当数据分析系统接收到用户对关系合并控件的触发时,可以显示下一级页面,该页面包括多个节点名称、可输入关系合并条件的输入框和一个合并控件。

[0124] 当数据分析系统接收到用户选中具有多个关系的两个节点名称、选择关系合并条件并触发合并控件时,将选中的两个节点合之间的关系并为一个,生成合并后的用户行为路径知识图谱。系统提供按标签分组合并、按属性值分组合并和全部合并这三种关系合并条件供用户选择。

[0125] 可能的实现方式五,在查询入口获取到关系拆分指令时,将合并后的关系重新恢复为可查看明细关系信息的状态,或者将合并后的关系以不同于合并时所采用的方式进行拆分,得到分析结果,分析结果包括关系拆分后的用户行为路径指示图谱。

[0126] 如图9所示,数据分析系统接收到用户对关系拆分控件的触发时,对合并后的节点之间的关系进行拆分。

[0127] 关系拆分为两种,一是数据分析系统接收到用户选择某个节点名称并触发关系拆分控件时,将合并后的关系重新恢复为可查看明细关系信息的状态的操作;二是数据分析系统接收到用户选择某个节点名称、输入关系拆分条件并触发拆分控件时,形成一些信息更丰富的不落盘图数据库的虚拟关系,具体指不以合并关系时使用的条件来拆分该关系,而是采用一种新的条件来拆分该关系,拆分后的关系是一种临时关系,并不会记录至图谱数据库。例如,节点A与节点B之间的关系是采用关系标签来进行合并的,在拆分该合并后的关系时,采用关系属性值对该关系进行拆分,得到的拆分结果只是临时展示,如果数据分析系统接收到刷新指令,该拆分结果就会消失,原有的合并后的关系仍然保留。

[0128] 可能的实现方式六,在查询入口获取到按关系属性筛选条件过滤关系指令时,隐藏不符合关系属性筛选条件的节点和对应的关系,得到分析结果,分析结果包括隐藏部分节点的用户行为路径知识图谱。

[0129] 数据分析系统接收到用户对按条件过滤关系控件的触发后,显示下级界面(图中未示出),该下级界面包括一个可输入关系属性筛选条件的输入框和一个过滤控件。

[0130] 数据分析系统在接收到用户在该输入框中输入关系属性筛选条件,并触发过滤控件后,隐藏不符合条件关系。在关系被隐藏后,可能出现一些孤立节点,数据分析系统在关系隐藏后同时隐藏该孤立节点。例如,节点A、节点B、节点C之间都存在关系,输入的关系筛选条件为隐藏节点A与节点B之间的关系和隐藏节点B与节点C之间的关系。此时,节点C将是一个孤立的节点,数据分析系统将会同时将节点C隐藏。

[0131] 可能的实现方式七,在所述查询入口获取到所述关系权重计算指令时,计算某一路径内部数量占两点间全部路径数量的比重,并按照所述比重对所述关系进行粗细标记,所述比重越大,所述关系之间的线条越粗,得到所述分析结果,所述分析结果包括对关系标注后的用户行为路径指示图谱。

[0132] 如图10所示,数据分析系统接收到用户对关系权重计算控件的触发后,会显示下一级页面,包含多个节点名称控件、函数选择控件及计算控件。

[0133] 数据分析系统在接收到用户选择存在多条路径的两个节点、需要使用的函数,并触发计算控件后,数据分析系统自动计算某一路径内部数量占两点间全部路径数量的比重,并按照比重自动对关系粗细进行标记,比重越大,关系线条越粗。数据分析系统对线条粗细的控制提供了对数归一化、MapMinMax归一化等方法,使粗细比例的显示相对合理。例如,节点A与节点F之间存在多条路径,经过权重计算后,路径1中包含5个节点,路径2包含3个节点,节点A与节点F之间的路径1的线条就会比路径2的线条粗。

[0134] 可能的实现方式八,在所述查询入口获取到所述图指标计算指令时,对所述用户行为路径知识图谱进行计算,得到链路文件,分析结果包括链路文件。

[0135] 当数据分析接收到用户对图指标计算控件的触发时,数据分析系统提供的固定算法对用户行为路径知识图谱进行整体分析,获得对应的链路文件。系统提供的固定算法包括环路识别算法、社区发现算法等图计算算法。其中链路指的是从一个节点到相邻结节点的一段物理线路,中间没有任何其他的交换节点。链路文件为多个链路的集合。

[0136] 例如,可以使用环路识别算法计算节点A与节点F在整个用户行为路径知识图谱中的所有路径,形成节点A到节点F的链路文件。

[0137] 可选的,查询指令还包括:节点统计指令和关系统计指令中的至少一项;在查询入口获取到查询指令时,根据查询指令对用户行为路径知识图谱中的内容进行分析,得到分析结果。

[0138] 可选的,在查询入口获取到查询指令为节点统计指令时,根据查询指令对用户行为路径知识图谱中的内容进行分析,得到分析结果可以包括下述实现:

[0139] 可能的实现方式一,在查询入口获取到节点统计指令时,根据节点标签或节点属性的名称统计所述用户行为路径知识图谱中各节点在所述用户行为路径知识图谱中的数量或本次查询出的数量。

[0140] 如图11所示,数据分析系统接收到用户触发节点统控件时,显示下一级界面,该界面包括多个可供用户选择节点统计条件的输入框、全图统计控件、取消控件以及确定控件。

[0141] 数据分析系统可接收用户选择节点标签的名称或节点属性的名称统计用户行为路径知识图谱中各节点在用户行为路径知识图谱中的数量或本次查询出的数量;本次查询出的数量指,数据分析系统接收到用户对节点统计控件的触发后,以当前时间为节点,可以显示的最大数量,该最大数量可以根据系统所处的终端设备的配置来调整。

[0142] 例如,如图11所示,数据分析系统接收到用户对全图统计控件的触发时,统计用户行为路径知识图谱中各节点在用户行为路径知识图谱中的数量;数据分析系统接收到用户对确定控件的触发时,统计本次查询的数量。

[0143] 可能的实现方式二,根据节点标签或节点属性的值,统计用户行为路径知识图谱中各节点在用户行为路径知识图谱中的分布情况或本次查询结果中不同值的分布状况。

[0144] 数据分析系统可接收用户在图11所示界面中选择节点标签或节点属性的值,统计用户行为路径知识图谱中各节点在用户行为路径知识图谱中的分布情况或本次查询结果中不同值的分布状况。

[0145] 可能的实现方式三,根据目标节点中某个具体的属性,统计所述目标节点在所述用户行为路径知识图谱中的值的分布情况或本次查询结果中值的分布情况。

[0146] 数据分析系统可接收用户在图11所示界面中选择目标节点中某个具体的属性,统计所述目标节点在所述用户行为路径知识图谱中的值的分布情况或本次查询结果中值的分布情况。

[0147] 可能的实现方式四,根据节点标签所属的功能模块划分多个社群,统计不同社群的内部的节点分布状况和对应的数量信息。

[0148] 数据分析系统可接收用户在图11所示界面中选择节点标签所属的功能模块进行社群划分,再统计各社群的内部节点分布状况和对应的数量信息。例如选择不同点击对象

的功能将节点标签划分为不同的节点社群后,再统计其中如入口和频道点击群体中用户访问各个入口的分布状况和数量信息。

[0149] 可选的,在查询入口获取到查询指令为关系统计指令时,根据查询指令对用户行为路径知识图谱中的内容进行分析,得到分析结果可以包括下述实现:

[0150] 与图11类似,数据分析系统接收到用户触发关系统控件时,显示下一级界面,该界面包括多个可供用户选择关系统计条件的输入框、全图统计控件、取消控件以及确定控件。具体的分析步骤与节点统计中所示步骤相同或相似,下述实现中不在赘述。

[0151] 可能的实现方式一,在查询入口获取到关系统计指令时,根据关系标签或关系属性的名称,统计用户行为路径知识图谱中各节点在用户行为路径知识图谱中的数量或本次查询出的数量。

[0152] 可能的实现方式二,根据关系标签或关系属性的值,统计所述用户行为路径知识图谱中各节点在用户行为路径知识图谱中的分布情况或本次查询结果中不同值的分布状况。

[0153] 可能的实现方式三,根据目标关系中某个具体的属性,统计所述目标关系在用户行为路径知识图谱中的值的分布情况或本次查询结果中值的分布情况。

[0154] 可能的实现方式四,根据拉依达 (3σ) 准则,统计主流路径和小众路径的集合和分布数量,按照全量显示的原则,突出显示全部的非非法路径。

[0155] 其中,主流路径为路径通过频率大于或等于频率阈值的路径;小众路径为路径通过频率小于所述频率阈值的路径;该阈值可以为数据分析系统用户通过对用户行为路径分析人为设置的。非法路径指超出数据分析系统程序允许范围的用户操作路径,例如,开发人员忘记将产品的旧代码旧入口屏蔽,某些用户通过了本不该开放的入口进入了某些页面,形成了非法路径;又或者是用户通过某些途径,恶意的获取到了后台直接的get或post请求,在直接发请求获取产品信息的过程中,日志记录到了这些行为,按时间顺序连接这些行为后,也能形成非法路径。

[0156] 可能的实现方式五,统计一定条件下所有的逆向路径。

[0157] 数据分析系统可接收用户选择统计一定条件下所有的逆向路径。例如,用户从一级页面进入二级页面,再从二级页面进入三级页面为正常路径,用户三级页面返回二级页面,再从二级页面返回一级页面,称为逆向路径。

[0158] 上述内容对数据分析系统获取到查询指令时,根据查询指令对用户行为路径知识图谱中的内容进行分析,得到分析结果,进行了说明。在此基础上,数据分析系统还可以接收用户对分析结果的进一步分析。

[0159] 可选的,在数据分析系统获得查询指令对应的分析结果后,数据分析系统可以接收用户对分析结果的进一步分析还包括下述实现:

[0160] 可能的实现方式一,数据分析系统还提供对分析结果的标记,包括,重点节点标记和重点关系标记。标记的内容不受拆分、合并和按条件过滤等操作的影响。数据分析系统只有再次接收到用户触发重点节点标记控件时,才可取消相应标记。

[0161] 可能的实现方式二,数据分析系统还提供对分析结果内容的输出下载。数据分析系统在接收到用户触发下载控件时,将目前分析结果完整的输出,避免了截图输出时的模糊和边界外信息丢失的问题。

[0162] 可选的,数据分析系统显示的第一界面中还包括滑动时间窗控件1201,如图12所示。

[0163] 数据分析系统在接收到用户的登录操作后,显示第一界面,第一界面显示的用户行为路径知识图谱默认会选中最长时间窗口进行全量的数据展示,也就是以用户进入数据分析系统的时间为节点,显示系统所支持的一次查询可以显示的最大数量。在数据分析系统接收到当用户对滑动时间窗控件1201的起止时间的修改后,可视化展示内容也会实时修改为只展示当前时间段内容的节点信息,相应的,关系信息也会随着节点信息的显示/隐藏而实时变动。

[0164] 可选的,数据分析系统显示第一界面还包括如下步骤:

[0165] 步骤1011、向图数据库发送post请求。

[0166] 数据分析系统向图数据库发送post请求,其中post请求为一种查询请求;图数据库储存有构建第一界面所需的底层数据,例如,用户在某些界面的点击次数、停留时间等。

[0167] 步骤1012、获取post请求的结果,请求结果为js对象简谱(javascript object notation,JSON)字符串数据。

[0168] 图数据库接收到数据分析系统的查询请求后,图数据库经过解析后向数据分析系统返回post请求结果对应的JSON字符串数据。

[0169] 步骤1013、将json字符串数据转化为可视化所需的json数据格式。

[0170] 步骤1014、基于预设的可视化渲染的技术框架、力导向布局、可视化样式配置信息以及对可视化所需的json数据格式的结果进行可视化渲染,得到可视化结果。

[0171] 具体的,基于预设的可视化渲染的技术框架,以力导向布局为基础,结合可视化样式配置信息,进行可视化渲染。渲染结束后呈现一个布局稳定且合理的可视化结果。可视化结果支持节点拖拽、节点关系的属性交互显示、右击节点后扩展等功能。

[0172] 可选的,如图13所示,步骤1014中可视化样式配置信息,包括:节点样式配置信息和关系样式配置信息。

[0173] 其中,节点样式配置信息包括:节点文字显示内容、节点大小、节点颜色、节点图标的一种或多种。

[0174] 其中,关系样式配置信息包括:关系文字显示内容、关系线条粗细、关系线条颜色、关系线条线型的一种或多种。

[0175] 示例性的,数据分析系统接收到用户触发可视化配置控件时,显示如13所示界面,数据分析系统接收到用户触发节点样式配置控件及节点文字控件时,显示可供用户输入文字的窗口。数据分析系统根据接收到用户输入的相关文字及触发保存配置控件时,配置相应的可视化界面。

[0176] 可选的,如图14所示,本申请实施例中的数据分析方法还包括如下步骤:

[0177] 步骤1401、收集用户行为日志。

[0178] 其中,用户行为日志包括:埋点日志和路由日志。

[0179] 步骤1402、将用户行为日志分类进行实体数据加工和关系数据加工,得到实体数据和关系数据。

[0180] 其中,实体数据加工包括:实体标签加工和实体属性加工;关系数据加工包括:关系标签加工和关系属性加工。

[0181] 可选的,实体标签加工包括:埋点实体标签加工和路由实体标签加工。

[0182] 其中,埋点的类型多样,埋点实体标签加工中按照不同的实体业务需求可分为不同的实体,包括但不限于下述的一种或多种:

[0183] 按照控件类型划分,可分为图片控件实体、文字控件实体、按钮控件实体等;按照不同点击对象的功能划分,可分为页签点击实体、入口和频道点击实体、具体内容点击实体、设置选项点击实体和其他功能按钮点击实体;按照业务功能划分,可分为支付流程实体、搜索流程实体。

[0184] 其中,路由实体标签,可划分为包括但不限于下述的一种或多种:

[0185] 按照页面深度划分,将不同的页面分为零级页面实体(首页实体)、一级页面实体,二级页面实体等;按照页面广度划分,设置不同的页面链接入度/出度数量阈值,将不同的页面划分为不同的实体,其中,阈值可以是若干个精确的整数(例如0,1,2……),也可以是若干个区间值域(例如[0,1],[2,4],[3,6]……);按照页面功能划分,将不同的页面分为信息收集页面、静态信息展示页面、动态信息查询页面。

[0186] 可选的,实体属性加工包括:将经过实体标签加工后得到的标签按照不同的业务需求划分为多种第一标签,将多种第一标签中的一种作为实体标签,以及将多种第一标签中除实体标签外的其余信息作为所述实体属性信息。例如,选择不同点击对象的功能,作为实体标签后,每个类别的实体就可以拥有控件类型、业务功能类型等属性。

[0187] 可选的,将单个用户的埋点或路由信息按时间顺序排列连接,即可得到该用户的原始用户行为路径关系数据;用户行为路径关系的类型多样,关系标签按照不同的关系业务需求可分为不同的关系,包括但不限于下述的一种或多种:

[0188] 按照路径通过频率划分,可分为主流路径、小众路径和非法路径;按照路径价值划分,可分为高价值路径,低价值路径和无价值路径等;按照路径驻留时间划分,可分为长时间驻留路径,短时间驻留路径,瞬时跳过路径等,也可分为多个驻留时间档次的路径;按照路径起始深度是否相同划分,可分为横向路径(深度相同)和纵向路径(深度不同);路径起始深度相同时,按照是否由浅向深访问,可分为顺序路径(源节点浅,目标节点深)和逆序路径(源节点深,目标节点浅)。

[0189] 具体的,主流路径为路径通过频率大于或等于频率阈值的路径;小众路径为路径通过频率小于所述频率阈值的路径;该阈值可以为数据分析系统用户通过对用户行为路径分析人为设置的。

[0190] 例如,可以将路径频率阈值设置为70%,所有用户行为路径中,用户通过该路径的频率大于70%时,该路径就为高价值路径;以购物产品为例,高价值路径,为较多用户通过此路径进行下单支付,

[0191] 低价值路径,为较少用户通过此路径进行下单支付,无价值路径,为几乎没有用户通过此路径进行下单支付。

[0192] 可选的,关系属性加工包括,将经过关系标签加工后得到的标签按照不同的业务需求划分为多种第二标签,将多种第二标签中的一种作为关系标签,以及将多种第二标签中除关系标签外的其余信息作为所述关系属性信息。例如选择路径通过频率,作为关系标签,每个类别的实体就可以拥有路径价值高低、路径驻留时长和路径方向等属性。

[0193] 步骤1403、基于实体数据和关系数据,构建用户行为路径知识图谱。

[0194] 具体的,利用实体数据和关系数据进行本体构建和知识映射,得到用户行为路径知识图谱。

[0195] 其中,本体构建包括:利用实体标签构建节点,实体标签与节点一一对应,各节点对应有各自的实体属性信息;利用关系标签构建关系,关系标签与关系一一对应,各关系对应有各自的关系属性信息。

[0196] 本体是一种广义的数据模型,仅对抽象出的一般类型进行建模,模型不包含具体实物信息,即本体用于定义实物类型和用于描述它们的属性。

[0197] 例如,选择埋点实体标签加工中的按控件类型划分中的图片控件实体、文字控件实体及按钮控件实体中的任一作为实体标签,构建对应的节点,选择其余信息如,业务功能、按钮功能等作为对应的属性信息。

[0198] 知识映射包括:根据本体构建中的实体标签、关系标签及实体标签和关系标签各自对应的属性信息,与实际数据处理中数据文件的字段进行一一映射。也就是为本体模型赋予真实世界的具体解释意义。例如,实体中某属性名称为“page_depth”(页面深度),实际数据源中某字段也为“page_depth”,那么这里就可以进行自动匹配。

[0199] 示例性的,图15为本申请实施例示出一种数据分析系统的构成示意图。如图15所示,数据分析系统包括分析模块和构建模块。分析模块为数据分析系统上层展示模块,用于接收到用户输入的查询指令后,根据查询指令对用户行为路径知识图谱中的内容进行分析,得到分析结果。构建模块为支持分析模块正常运行的底层模块。

[0200] 其中,分析模块包括:数据分析层。构建模块包括:日志收集层、数据处理层、图谱构建层以及图谱可视层。

[0201] 可选的,构建模块实现分析模块的正常运行包括下述步骤:

[0202] S1501、日志收集层收集用户行为日志。

[0203] 其中,用户行为日志包括:埋点日志和路由日志。

[0204] S1502、数据处理层将日志收集层收集的用户行为日志分类进行实体数据加工和关系数据加工,得到实体数据和关系数据。

[0205] 其中,实体数据加工包括:实体标签加工和实体属性加工;关系数据加工包括:关系标签加工和关系属性加工;实体数据加工与关系数据加工的过程与步骤1402相同,此处不在赘述。

[0206] S1503、图谱构建层基于数据分析层加工好的实体数据和关系数据,构建用户行为路径知识图谱。

[0207] 图谱构建层构建用户行为路径知识图谱的过程与步骤1403相同,此处不在赘述。

[0208] S1504、图谱可视层对图谱构建层构建的用户行为路径知识图谱进行可视化配置。

[0209] 图谱可视层对构建好的用户行为路径知识图谱进行可视化配置的方式与图13所示内容相同,此处不在赘述。

[0210] 如图16所示,本申请实施例还提供一种数据分析装置160,包括:

[0211] 第一显示模块1601,用于显示第一界面,第一界面包括查询入口以及用户行为路径知识图谱的部分或全部;用户行为路径知识图谱包括多个节点、多个节点之间的关系、任一节点的节点属性以及任一关系的关系属性;节点属性包括节点类型、点击次数的一种或多种,关系属性包括用户行为路径的路径频次、路径类型或路径发生时间的一种或多

种。

[0212] 分析模块1602,用于在查询入口接收到查询指令时,根据查询指令对用户行为路径知识图谱中的内容进行分析,得到分析结果。

[0213] 第二显示模块1603,用于显示第二界面,第二界面包括分析结果。

[0214] 本申请实施例还提供了一种电子设备,该电子设备通过系统总线连接的处理器和存储器。其中,该电子设备的处理器用于提供计算和控制能力。该电子设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该电子设备的数据库用于存储可视化数据、视图资源数据等。该计算机程序被处理器执行时以实现一种数据分析方法。

[0215] 本申请实施例还提供了一种计算机可读存储介质。上述实施例中描述的方法可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。如果在软件中实现,则功能可以作为一个或多个指令或代码存储在计算机可读介质上或者在计算机可读介质上传输。计算机可读介质可以包括计算机存储介质和通信介质,还可以包括任何可以将计算机程序从一个地方传送到另一个地方的介质。存储介质可以是可由计算机访问的任何目标介质。

[0216] 一种可能的实现方式中,计算机可读介质可以包括随机存取存储器(Random Access Memory, RAM),只读存储器(Read-Only Memory, ROM),只读光盘(compact disc read-only memory, CD-ROM)或其它光盘存储器,磁盘存储器或其它磁存储设备,或目标于承载的任何其它介质或以指令或数据结构的形式存储所需的程序代码,并且可由计算机访问。而且,任何连接被适当地称为计算机可读介质。例如,如果使用同轴电缆,光纤电缆,双绞线,数字用户线(Digital Subscriber Line, DSL)或无线技术(如红外,无线电和微波)从网站,服务器或其它远程源传输软件,则同轴电缆,光纤电缆,双绞线,DSL或诸如红外,无线电和微波之类的无线技术包括在介质的定义中。如本文所使用的磁盘和光盘包括光盘,激光盘,光盘,数字通用光盘(Digital Versatile Disc, DVD),软盘和蓝光盘,其中磁盘通常以磁性方式再现数据,而光盘利用激光光学地再现数据。上述的组合也应包括在计算机可读介质的范围内。

[0217] 本申请实施例是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理单元以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理单元执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0218] 以上的具体实施方式,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,以上仅为本发明的具体实施方式而已,并不用于限定本发明的保护范围,凡在本发明的技术方案的基础之上,所做的任何修改、等同替换、改进等,均应包括在本发明的保护范围之内。

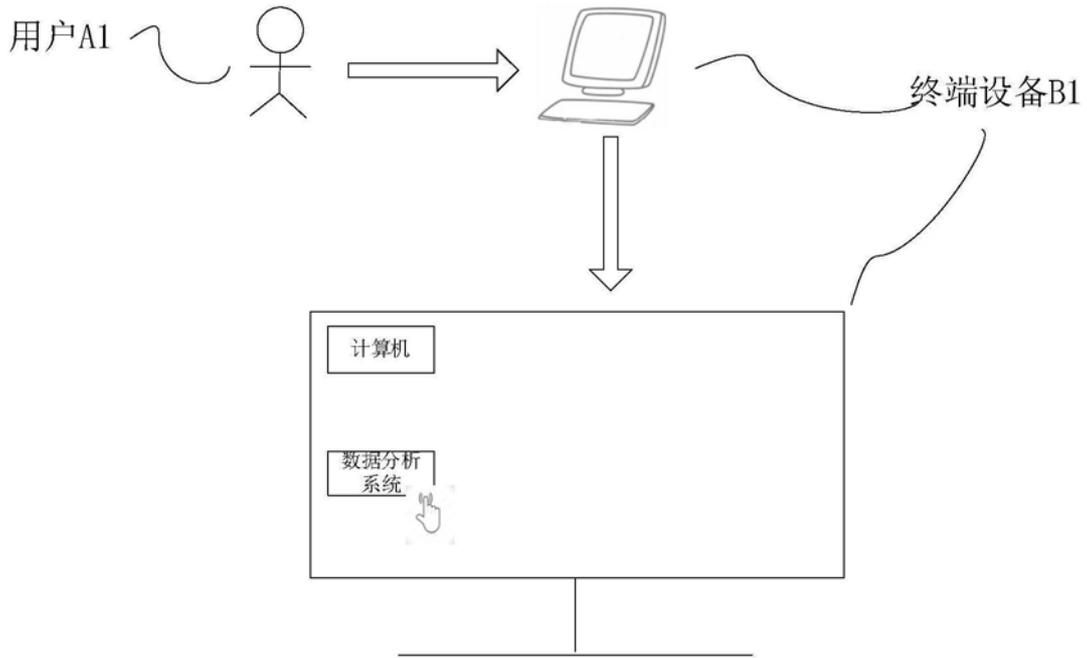


图1

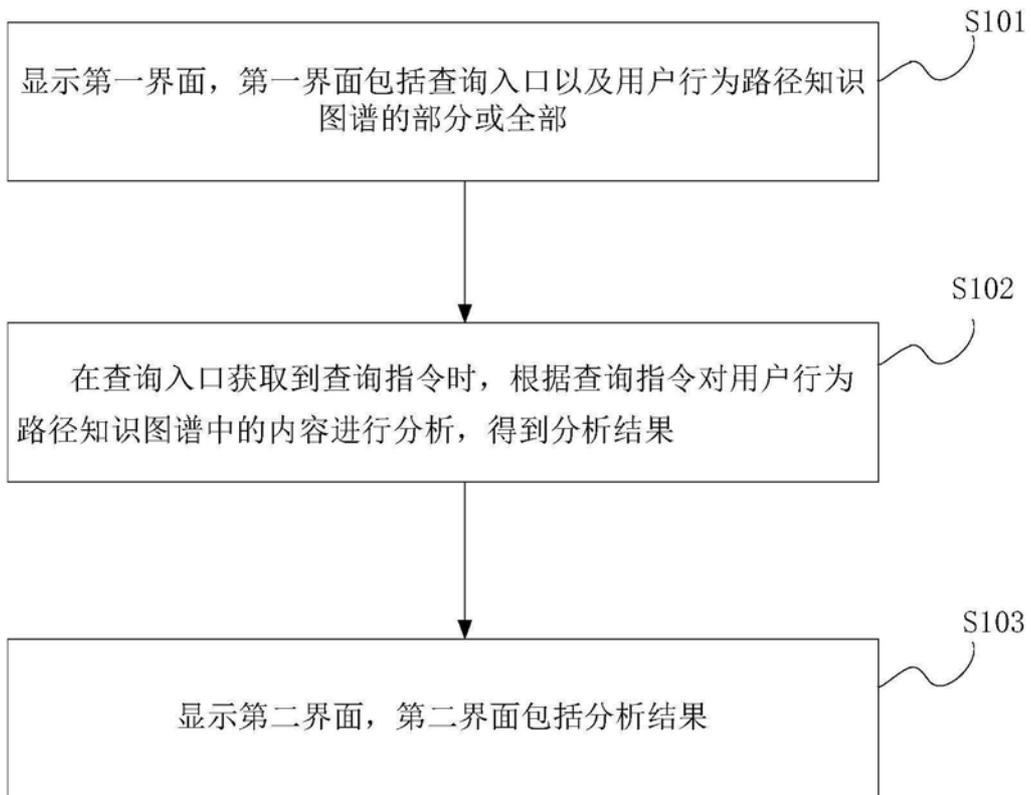


图2

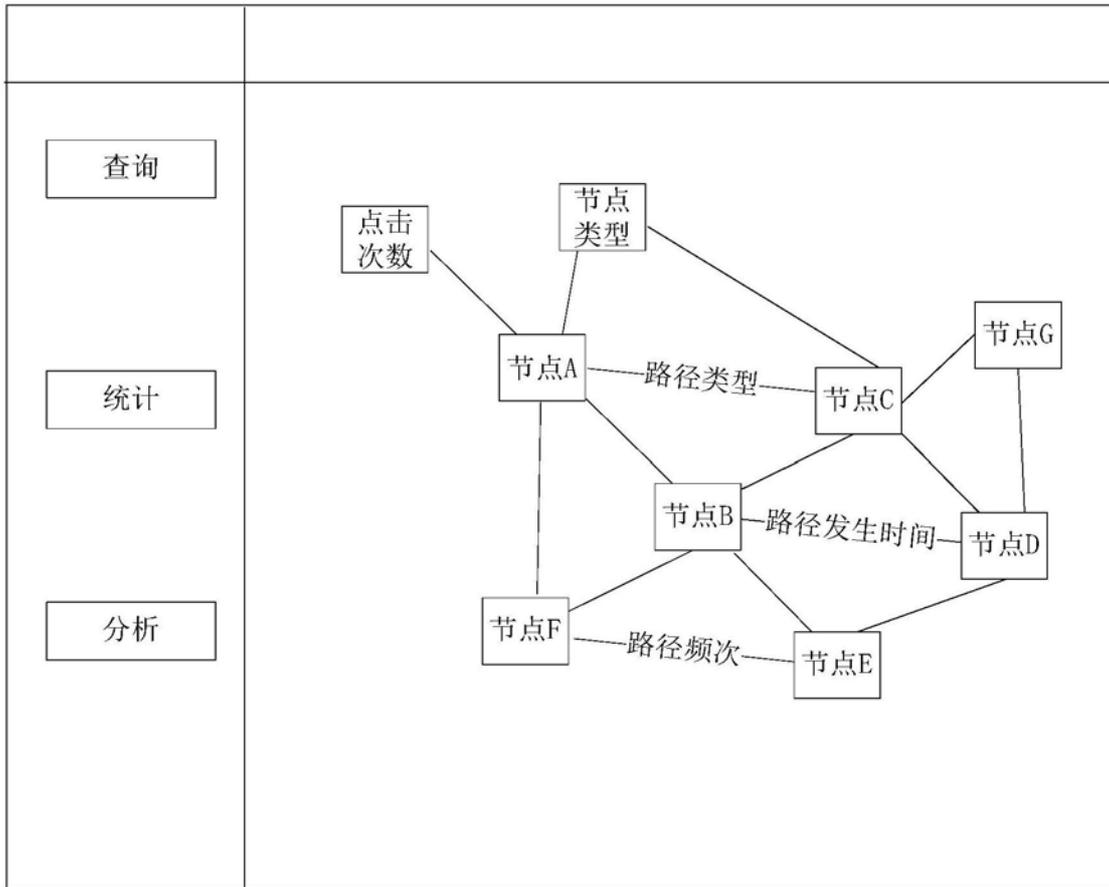
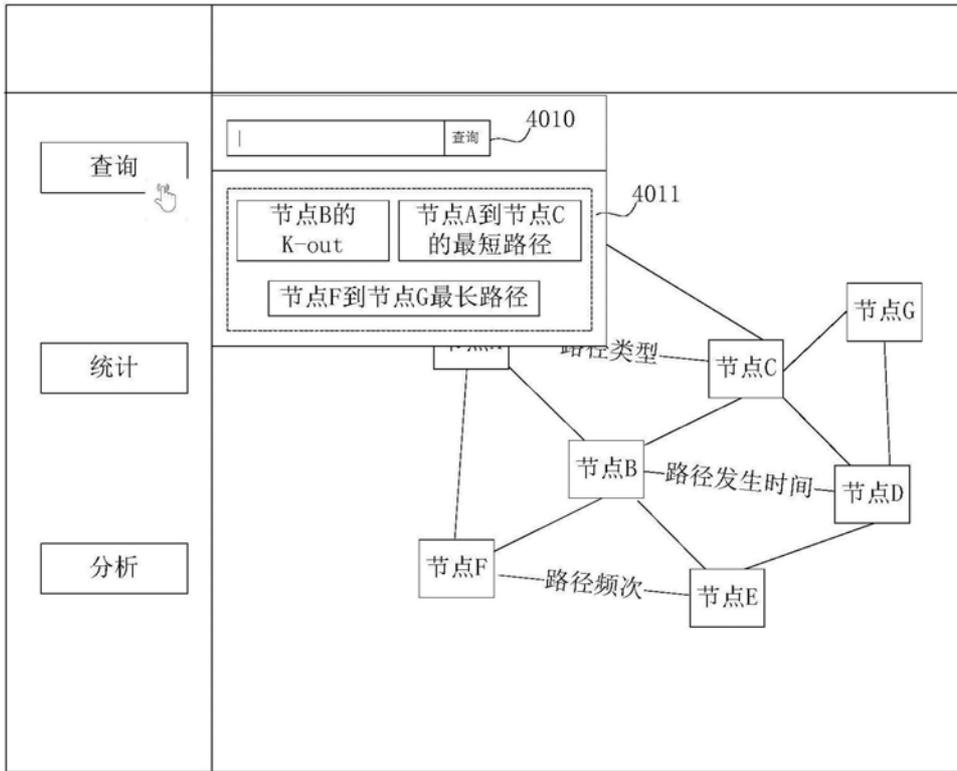
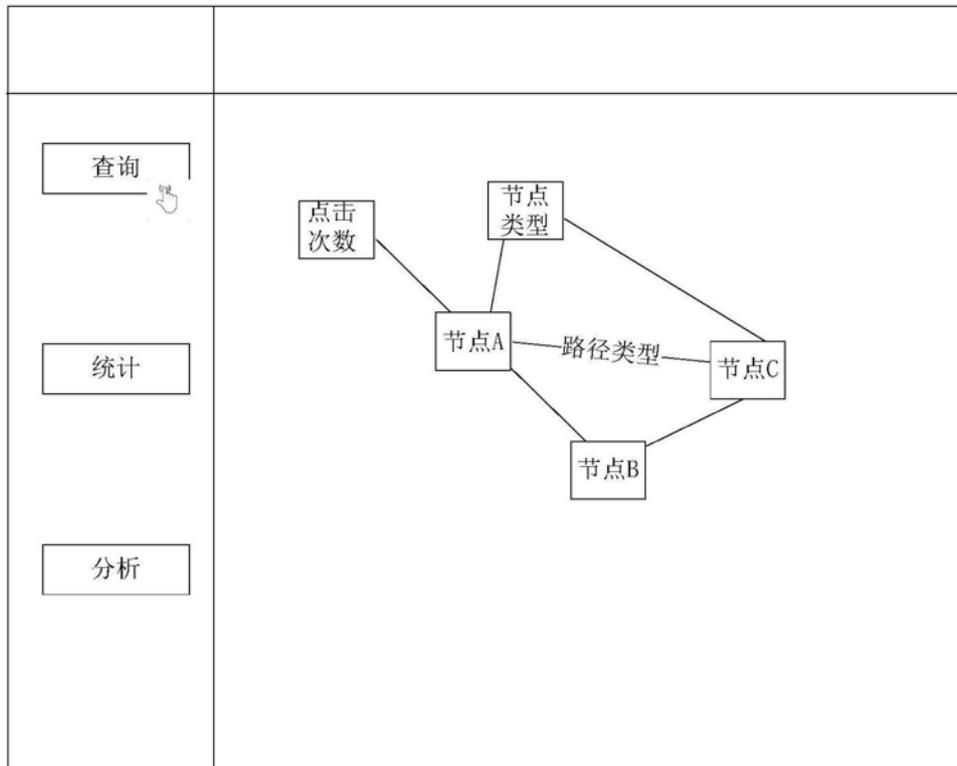


图3

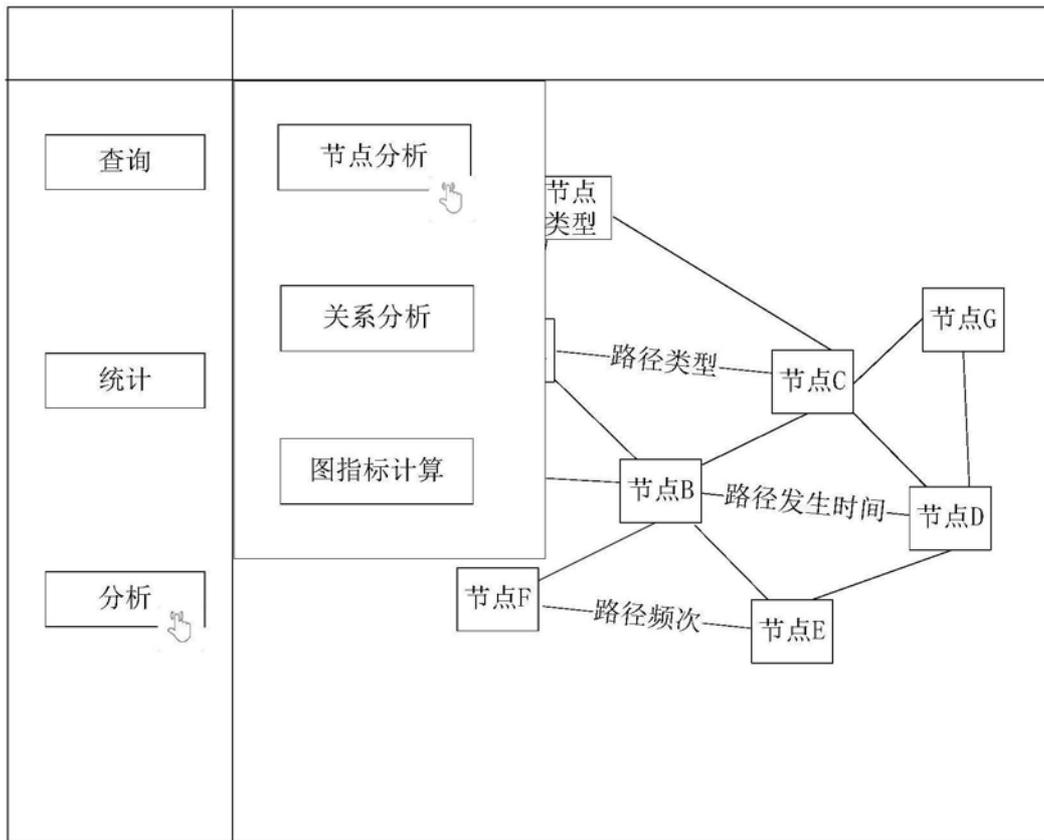


(a)

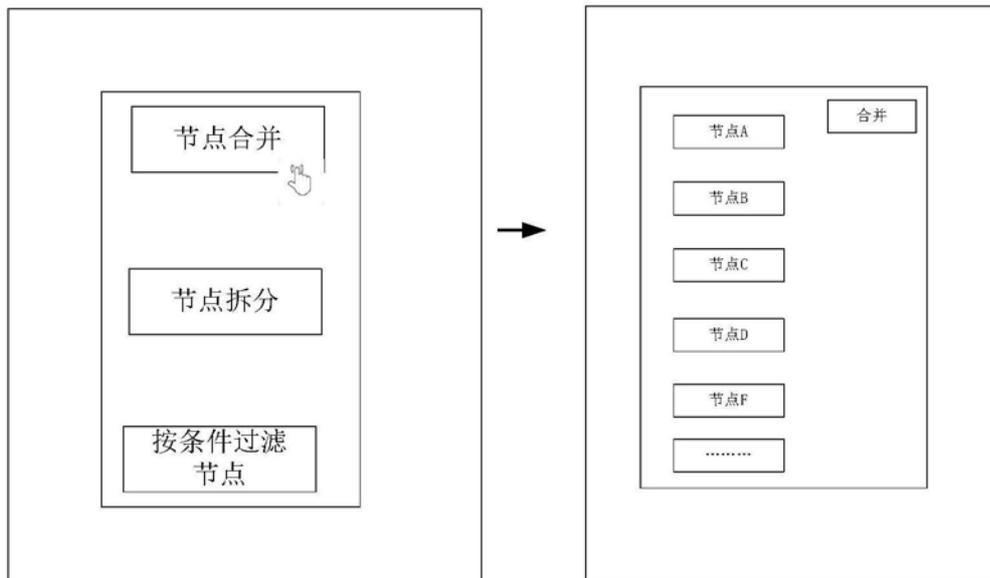


(b)

图4



(a)



(b)

(c)

图5

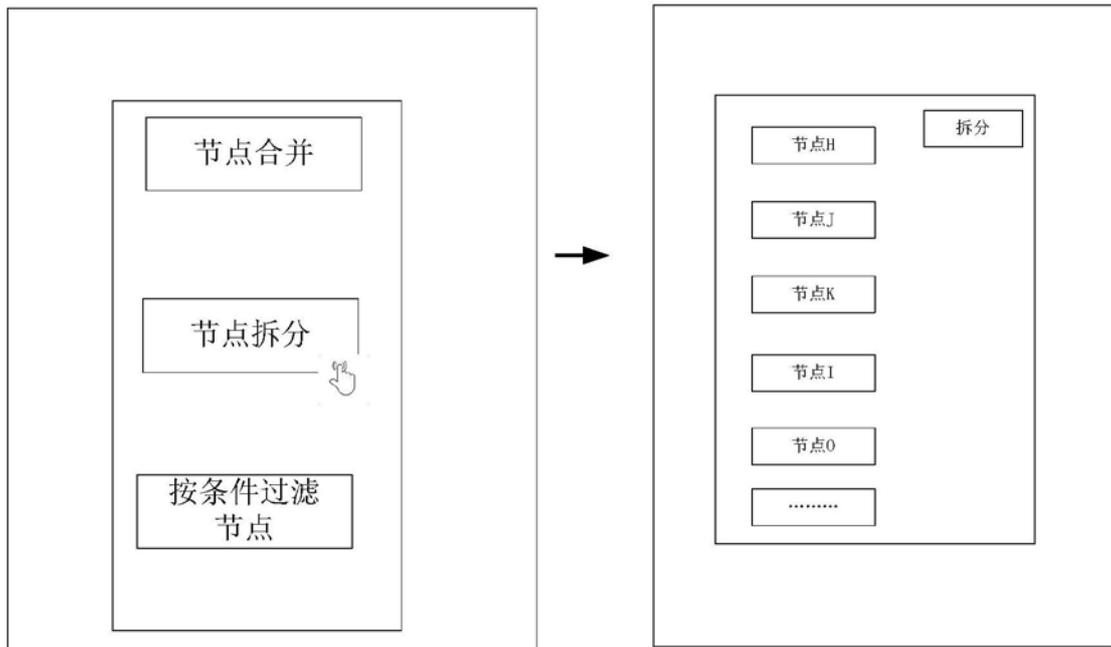


图6

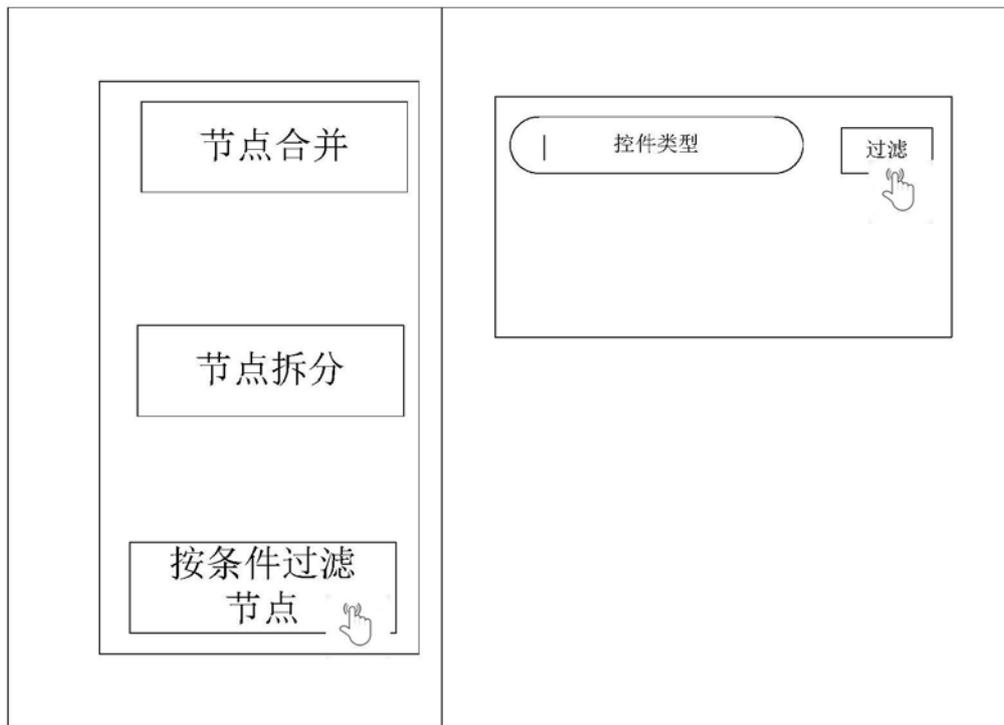


图7

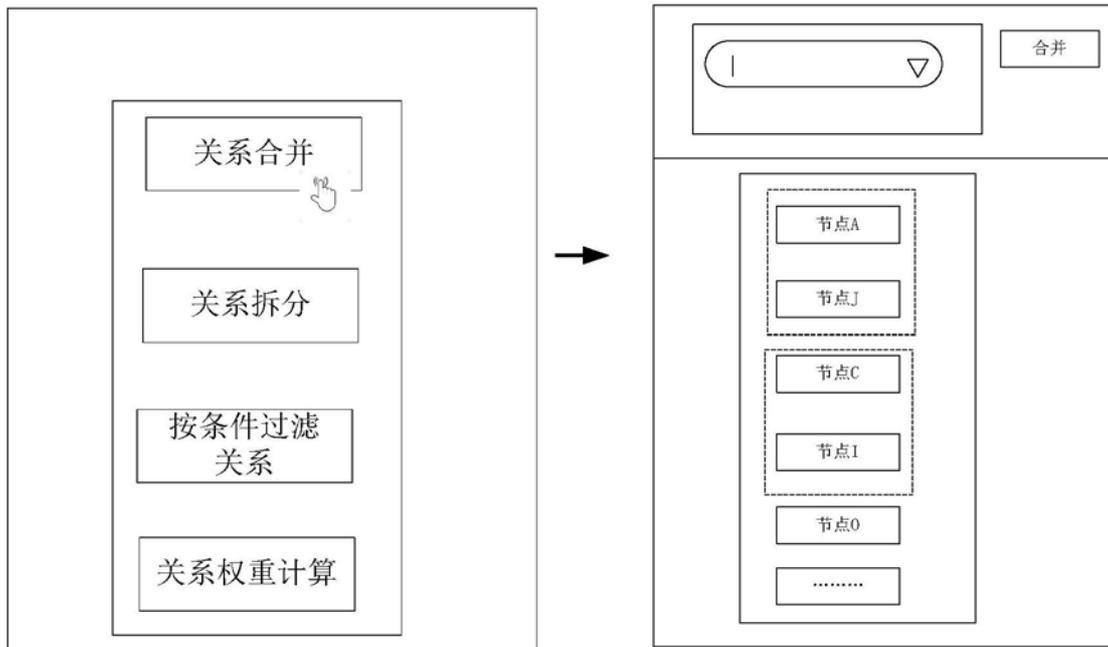


图8

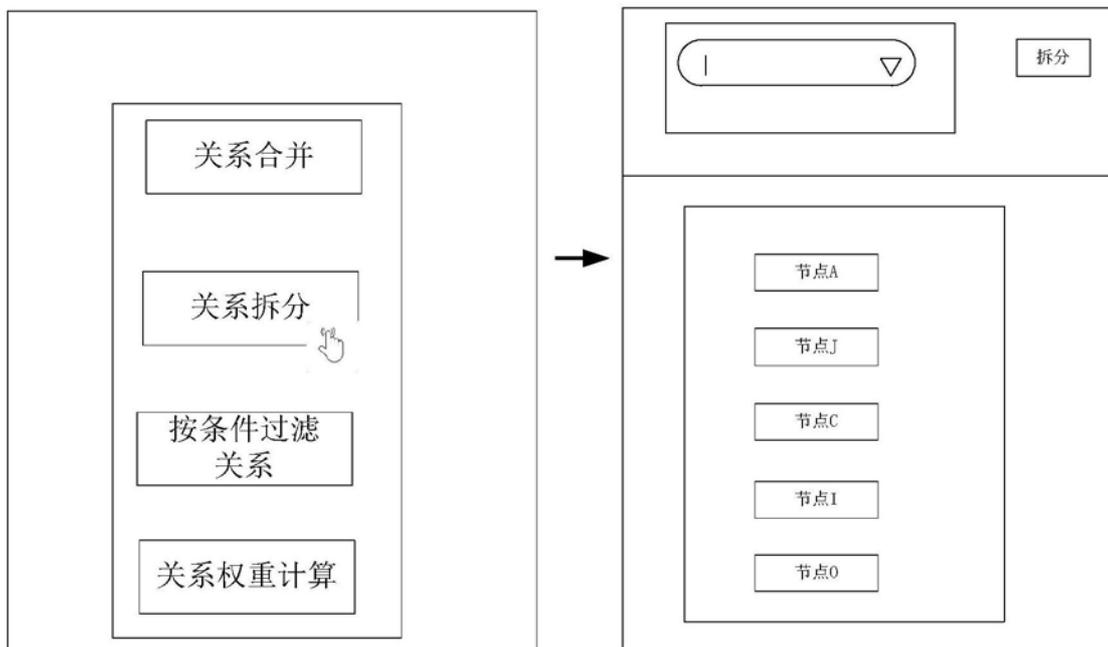


图9

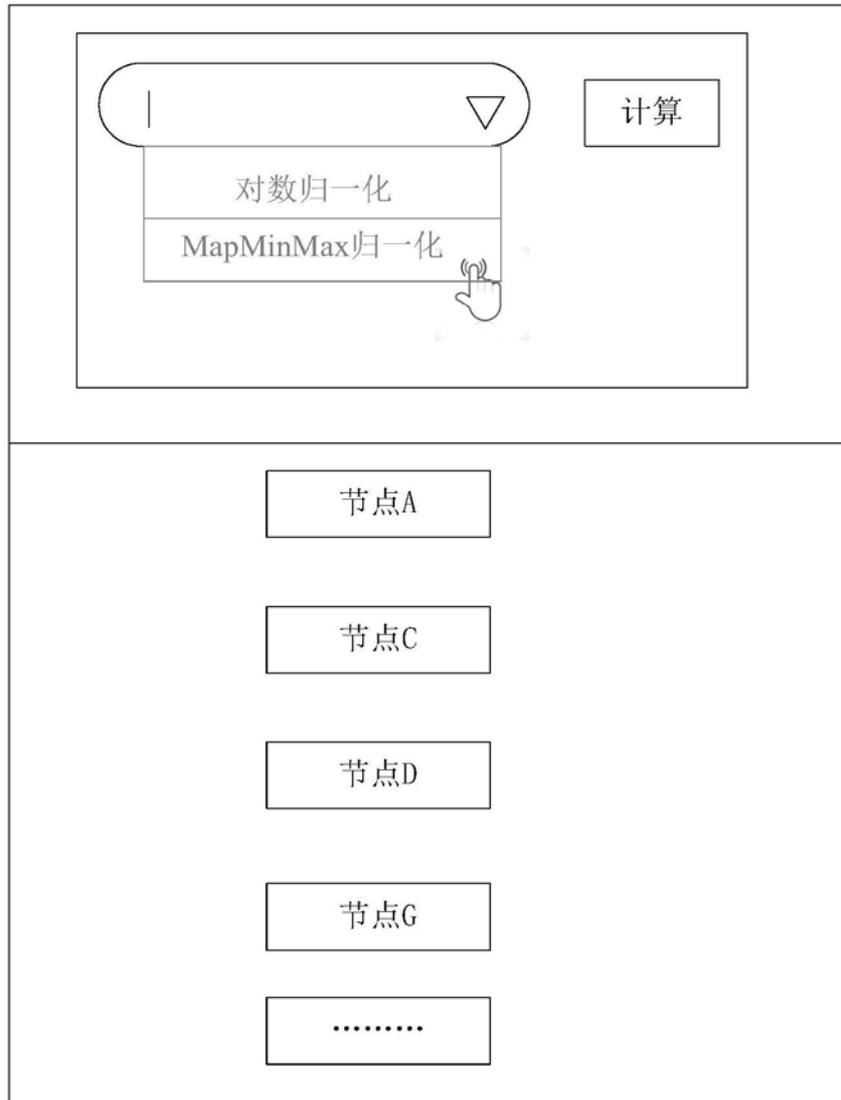
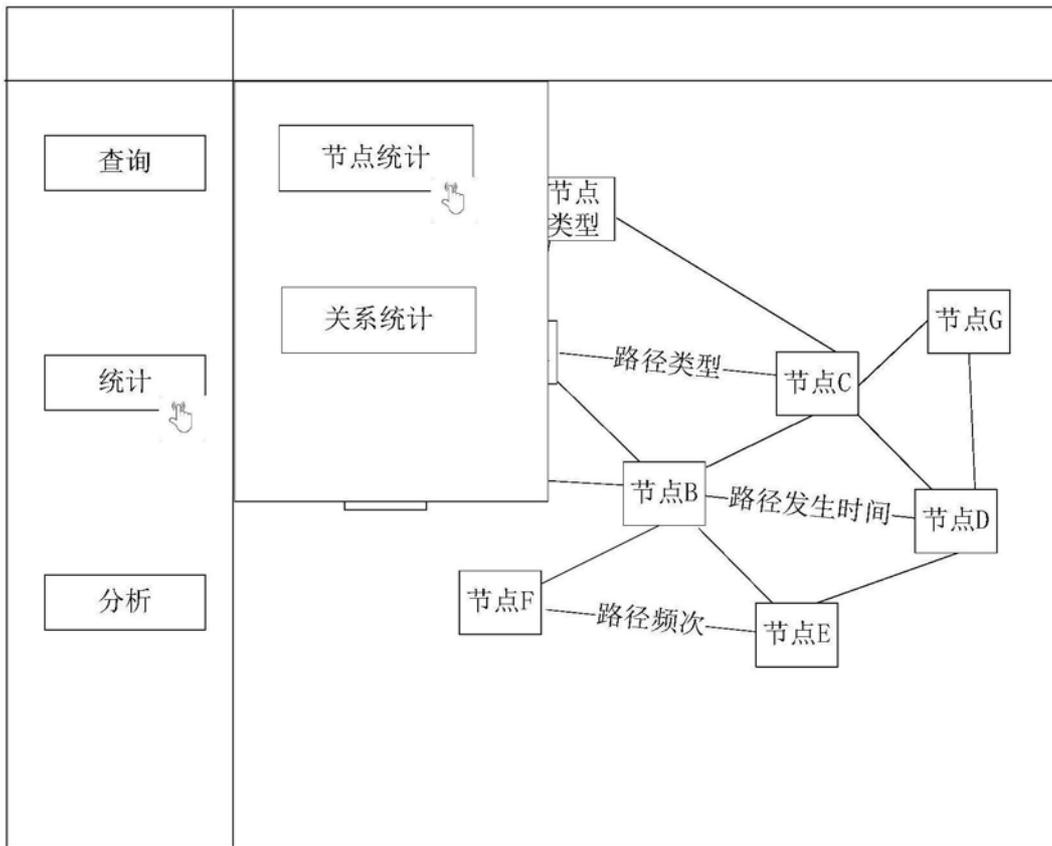


图10



The configuration dialog box is titled '选择统计规则' (Select Statistical Rules). It contains several dropdown menus for selecting nodes and components. The first row has three dropdowns: '节点A' (Node A), '控件类型' (Control Type), and an empty dropdown. The second row has three dropdowns: '一级页面' (First-level Page), an empty dropdown, and another empty dropdown. At the bottom left is a button labeled '全图统计' (Full Graph Statistics). At the bottom right are two buttons: '取消' (Cancel) and '确定' (Confirm).

图11

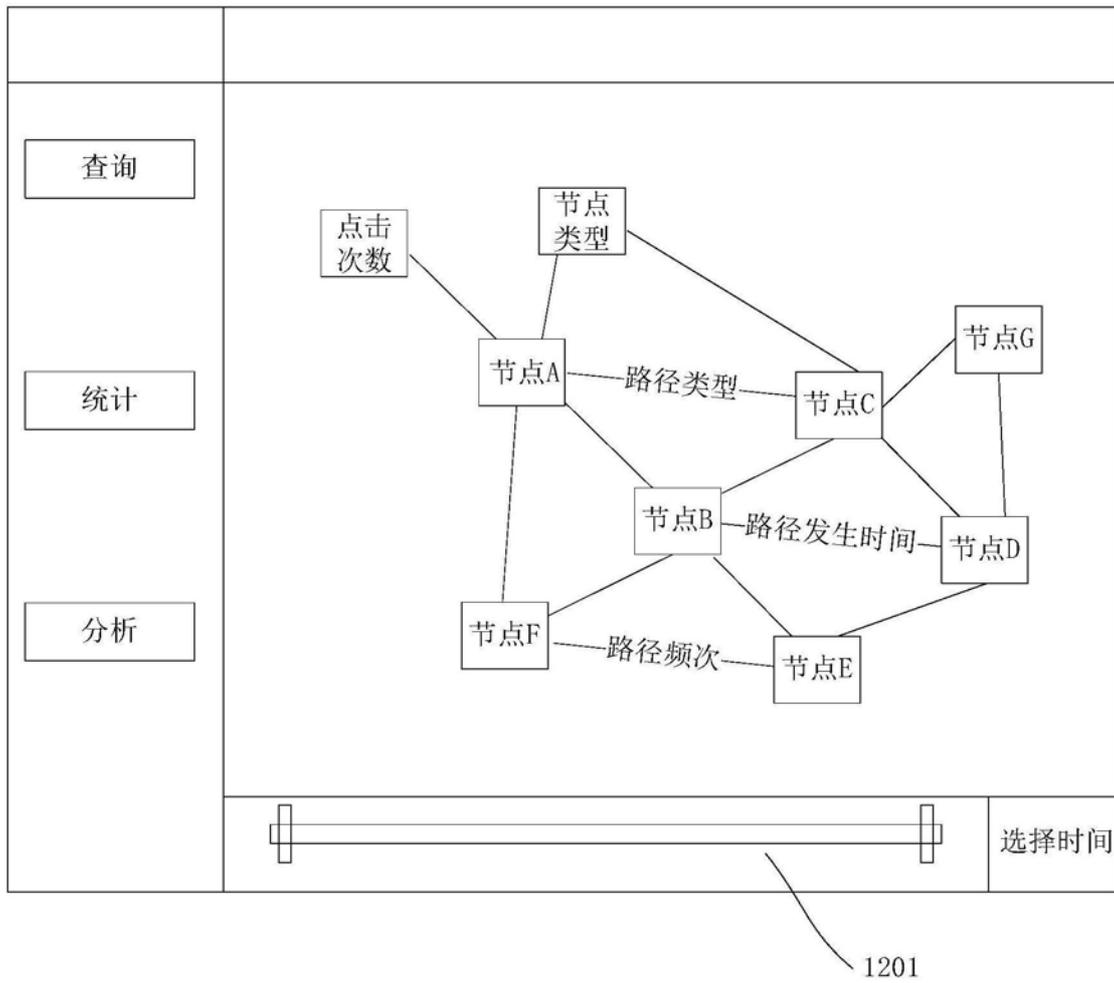


图12



图13

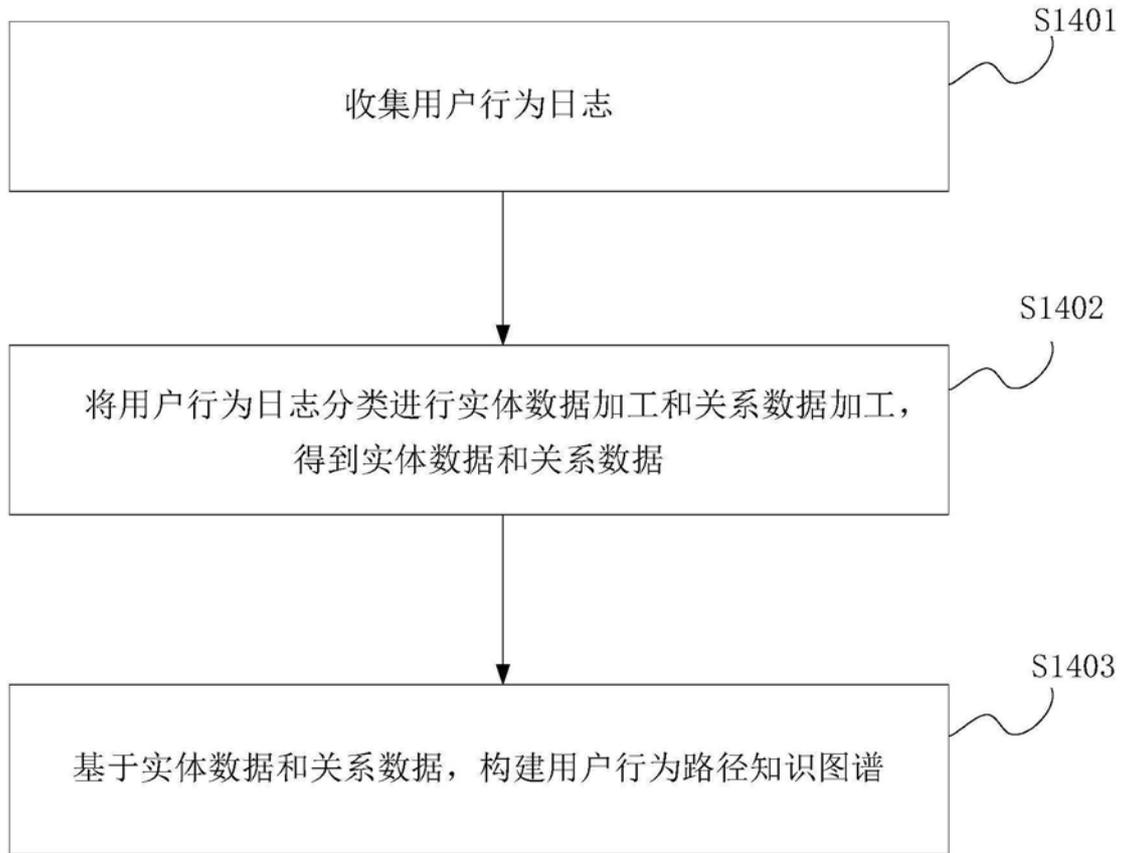


图14

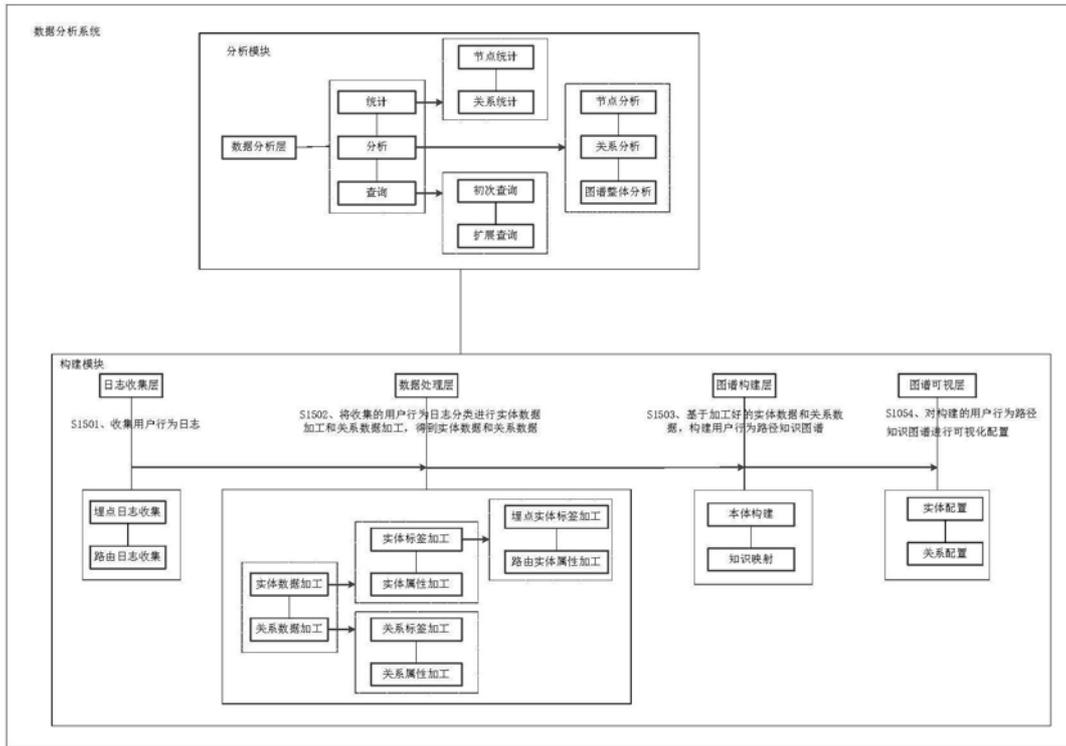


图15

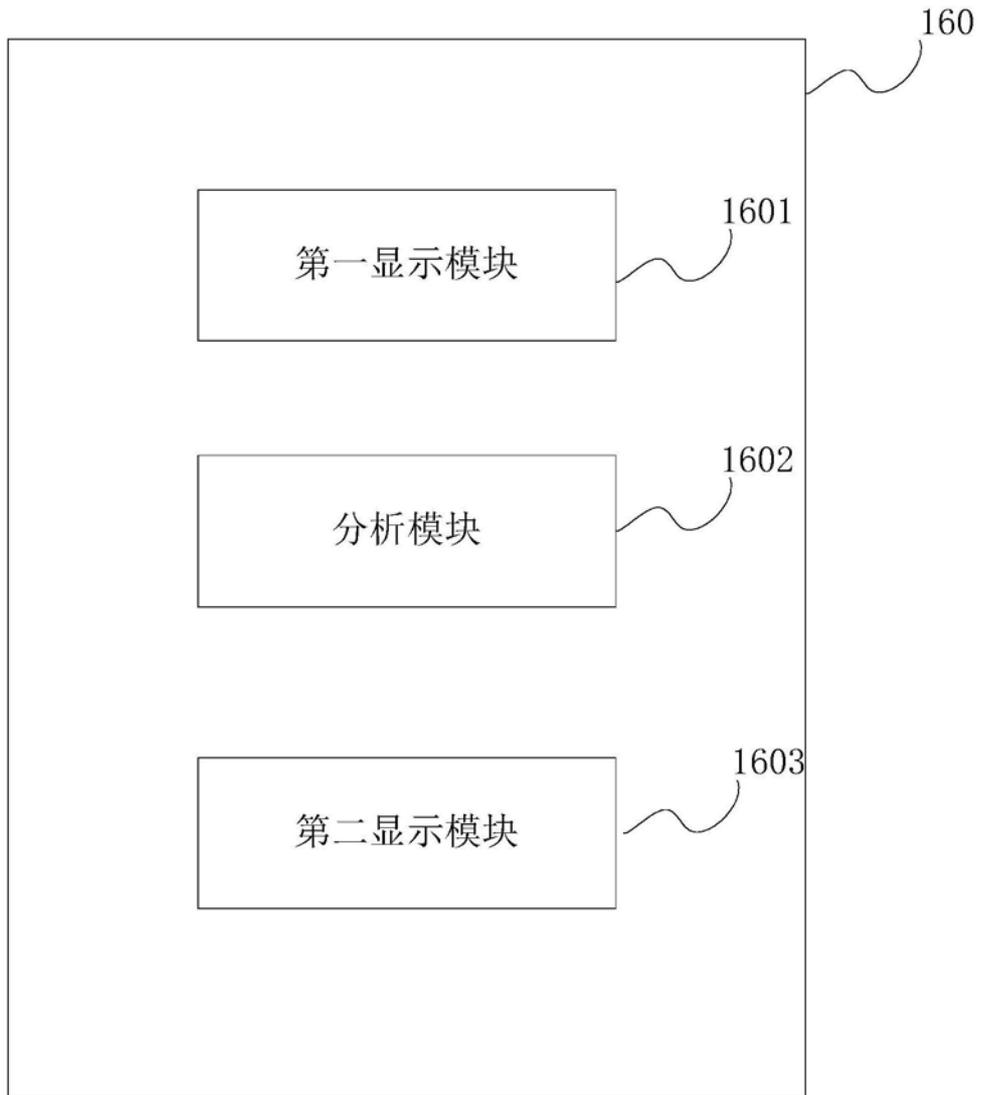


图16