

动态数据关联网络的表示及搜索方法^①



陈琼, 李良, 刘钊岐

(中电海康集团有限公司, 杭州 311100)

通信作者: 陈琼, E-mail: 22245087@qq.com

摘要: 设计和实现了一种动态数据关联网络的表示及搜索方法和系统, 能够在数据实体较多、关联关系较复杂时, 帮助用户获得实体周边关联关系, 并通过引导式交互不断动态扩展; 在已知可能有关联的多个实体时, 采用分布式计算最小连通图算法, 搜索出其关联网络. 应用实例表明, 本方法和系统能够取得很好的实际效果, 采用本方法的应用系统已经在智慧城市、平安城市、城域物联网等多个工程项目中获得落地应用.

关键词: 数据关联网络; 数据可视化; 关联关系搜索; 大数据

引用格式: 陈琼, 李良, 刘钊岐. 动态数据关联网络的表示及搜索方法. 计算机系统应用, 2022, 31(8): 115-124. <http://www.c-s-a.org.cn/1003-3254/8601.html>

Representation and Search Method of Dynamic Linked Data Network

CHEN Qiong, LI Liang, LIU Zhao-Qi

(CETHIK Group Co. Ltd., Hangzhou 311100, China)

Abstract: This study designs and develops a representation and search method and a relevant system for dynamic linked data networks. The method can help users obtain the association around an entity when there are complicated entities and associations, and the network can be dynamically expanded through guided interaction. When users know multiple entities which are suspected to be associated, the minimum connected graph algorithm based on distributed computing is used to search out the association network. Application cases show that the proposed method and system can achieve good results and user experiences, and the system based on the method has been applied in many engineering projects such as Smart City, Safe City and Metropolitan IoT.

Key words: linked data network; data visualization; association search; big data

在大数据时代, 人与人、人与物, 乃至物与物之间的关系日趋复杂多样, 如何从海量数据中挖掘出所关心的信息, 并且以友好、可理解的形式展现出来, 是很多应用场景非常关心的问题. 可视化技术起源于 20 世纪 80 年代出现的科学计算可视化 (visualization in scientific computing), 它是指利用计算机图形学、计算机图像处理、计算机信号处理等方法对数据、信息、知识的内在结构进行表达^[1]. 进入 21 世纪, 单一的可视化已不能满足人们日益增长的对于挖掘数据中存在的关联关系的需求, 可视化逐渐发展为一个涉及数据挖掘、人机交互、计算机图形学等的交叉学科^[2]. 在表现

人人、人物, 乃至物物之间的数据关联关系时, 数据关联网络是常见, 也是直观的可视化手段. 但是目前的数据关联网络表示方法, 一般都是静态的、一次性的, 适用于需要表现的实体较少, 关系较为简单的情形. 当实体较多, 之间的关联较为复杂时, 传统方法无论从后台的处理能力角度, 还是前端的用户体验角度, 都较难满足复杂的表达数据关联的需要. 并且, 从用户角度出发, 用户往往是在已知几个人、物的情况下, 希望得到这些人、物之间的直接或者间接的关联关系, 从而进一步得到围绕这些已知人、物的关联网络关系. 本文提出一种可动态扩展的引导式交互数据关联网络的表示及

^① 基金项目: 国家科技重大专项 (2018ZX01028102)

收稿时间: 2021-11-01; 修改时间: 2021-12-02; 采用时间: 2021-12-08; csa 在线出版时间: 2022-03-31

搜索方法和以此实现的系统,能够较好地解决这些问题.

1 概念和实现框架

1.1 数据关联的概念和理论

数据关联技术来自对语义网研究,演化的结果是数据网络.关联数据是指网络中发布和连接的有结构数据的一组最佳实践^[3].语义网中关联数据技术的出现,改变了传统的构造、发布、访问和集成数据的方式^[4],它采用资源描述框架(RDF)作为基本数据模型,制定了关于数据表示和链接的规范^[5].

“一图胜千言”这句谚语告诉我们:一张图传达的信息等同于相当多文字的堆积描述^[6].为更好地显示关系网络的结构特征和进行相关的计算,一般会使用图论相关的理论对关系网络建模,表示社会网络中人人、人物或者物物的拓扑结构和属性信息.国际上对社会关系网络的关联模型及其可视化研究是一个热点^[7-10].丁楠等^[11]提出了一种基于关联数据技术进行多源政府信息聚合的模型,徐长志等^[12]对关联数据的空间关联进行了探讨,王柳等^[13]对学者社交网络提出了一种关联模型,王晰巍等^[14]对社交网络舆情的图谱构建提出了一种方法,杜晓林^[15]则对大规模社会网络的可视化算法进行了研究.

社会关系网络具有聚类特征,社会网络的聚类系数一般远高于随机生成的网络的聚类系数值,从而形成所谓社团结构^[16].从图论的角度看,社团结构是网络的节点集合,该集合节点间的连接较为稠密,而其外节点的连接则较为稀疏.社会关系网络的这种聚类特征,是数据关联网络的动态扩展,在引导式交互中逐步发掘出周边有价值的关联信息的理论保证.

虽然社会关系网络异常复杂,但通常来说个人之间的关联具有“最小世界”特征.20世纪60年代,Travers等论证了“六度空间”或曰“最小世界”理论,即人们相互之间联系的距离大概为6个人^[17].六度空间是社会网络中“小世界”主要特征,说明在现代超大规模的社会关系网络中,人与人之间的平均最短距离远少于网络中成员的数量^[18].这种特征为在数据关联网络中进行代价有限的搜索提供有价值的信息提供了理论基础.

1.2 数据关联网络的技术实现框架

数据关联网络分析与可视化的实现框架如图1所示,可以分为4个阶段.

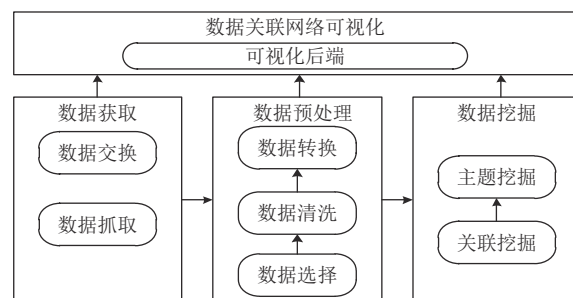


图1 数据关联网络实现框架

(1) 数据获取: 依据不同的应用场景,采用各类技术手段获得数据,例如在智慧城市场景,通常需要通过数据交换获得来自不同部门的数据,在如网络舆情场景,则常常使用抓取工具获得网上海量信息.

(2) 数据预处理: 对所关心的数据进行选择过滤,并对数据进行清洗,将数据转变成“干净”的数据,然后对数据格式进行转化,变成统一的格式.

(3) 数据关联挖掘: 使用合适的算法,针对不同的领域主题进行规则下的挖掘,使得数据具有面向不同主题的结构.

(4) 可视化: 针对不同应用和不同主题,采用适当的技术和可视化形式,对数据集进行可视化;对关联规则结果的可视化,可以使得数据关联分析的结果直观易懂,让用户能够对可视化结果做出解释和评估,从而形成有价值的信息和知识.

2 系统的设计与实现

2.1 动态数据关联网络表示方法的系统设计

针对前述对社会网络等应用场景的可视化需求,本文提出一种数据关联网络图的表示方法和系统,使得用户在已知某个数据(通常是某个人、物等实体)的情况下,能够通过引导式交互互动的办法,逐步扩展关联网络图,利用前述社会关系网络的社团性聚类特征,通过有限的步骤,挖掘出围绕该数据实体周边的有价值信息.

本系统通过后端建立数据关联网络模型,响应来自前端的请求并返回结果;前端根据用户的操作,生成操作命令并向后端请求数据,并在收到后端返回数据后根据数据关联网络模型绘制和更新数据关联网络图;该数据关联网络图在用户的操作下能够不断动态扩展.本方法及系统理论上能够支持无限多元素的实体关系网络,并且能够扩展不同类型的实体、关系和实体聚

合.本方法和系统能够非常直观、友好地展现复杂的数据关联图谱,极大地方便数据分析人员工作.

数据关联网络表示系统前后端交互流程如图2.

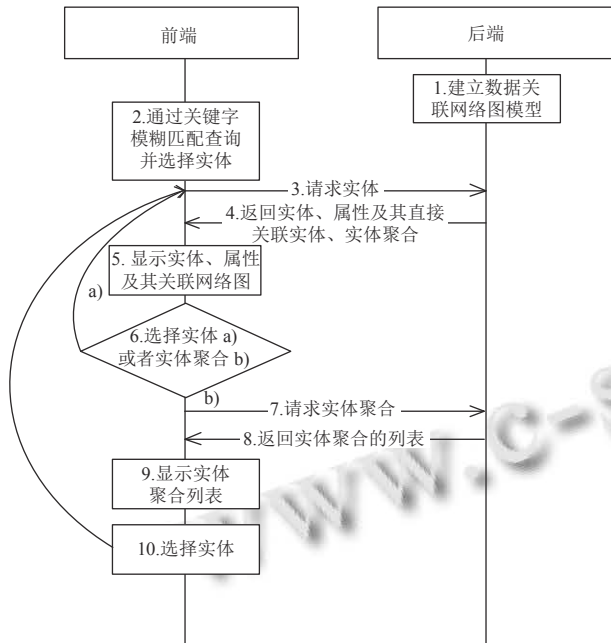


图2 动态扩展的前后端交互流程

(1) 系统后端建立数据关联网络图模型,模型的元素包括实体、关系和实体聚合;

(2) 用户选择实体类型并通过关键字模糊匹配查询,在系统后端返回的实体候选列表中选择实体;

(3) 系统前端将所选择的实体发给后端;

(4) 系统后端根据数据关联网络图模型,将该实体直接关联的实体、关系和实体聚合返回给前端,同时也返回该实体的属性;

(5) 系统前端显示该实体及其属性,并根据数据关联网络图模型生成和该实体直接关联的实体、关系和实体聚合的网络图,同时更新所返回的新增实体与现有实体的关联网络连接;

(6) 用户点击构成当前数据关联网络图中可选中的任何元素,可选中的实体包括实体或者实体聚合;

a) 如选中的为实体,则执行步骤(3)–(6);

b) 如选中的为实体聚合,则执行步骤(7)–(10).

(7) 系统前端将所选择的实体聚合发给后端;

(8) 系统后端将构成该实体聚合实体列表返回前端;

(9) 前端显示该实体聚合的实体组成列表;

(10) 用户选择实体列表中某个实体,执行步骤(3)–(6);

用户可反复执行步骤(3)–(10),则数据关联网络图不断动态扩展.

在前述的数据关联网络图模型,元素包括实体,关系和实体聚合;其中实体可以包括普通属性,也可以包含列表属性.在前述的关系表示实体之间的有向关系,这种关系可以单向,也可以双向.在前述的实体聚合用于在同类实体较多时,将多个同类实体表示为一个实体聚合.在前述的步骤(5),前端将当前选中的实体聚焦,并显示其普通属性和列表属性.在显示与当前选中实体直接关联的实体时,如果存在同类实体,则显示实体聚合.在前述的步骤(9),前端将当前选中的实体聚合聚焦,并显示其所组成的实体列表供选择.如果实体聚合组成的实体个数小于等于 N (N 可定义为2或者其他值),则展开该实体聚合包括的所有实体列表.在前述的数据关联网络图可以根据用户的点击交互不断扩展,数据关联网络图会在扩展后自动平衡到最佳位置;在前述的数据关联网络图中的可选中元素可以自由拖动,数据关联网络图会在拖动后自动平衡到最佳位置.

2.2 动态数据关联网络搜索方法的系统设计

上述数据关联网络的表示方法可以不断动态扩展,但用户更需要提供一种搜索方法,在已知两个或者两个以上的实体的情况下,希望得到这几个已知实体的关系.根据前述的“最小世界”理论,这几个怀疑有关联的实体,可通过代价有限的搜索步骤得到其关联关系.通过后端建立数据关联网络模型,采用分布式计算扩展关节点、并计算最小连通图算法,计算搜索端点之间的关联关系,响应来自前端的搜索请求并返回结果;前端根据后端返回数据生成数据关联网络图,其中包括搜索端点.本方法为用户在知道两个或多个实体,并且想知道它们之间直接或者间接关系时,提供有效的搜索手段,并且通过直观、友好的方式展现出来.

本搜索系统的前后端交互流程如图3所示.

(1) 系统后端建立数据关联网络图模型;

(2) 用户选择实体类型并通过关键字模糊匹配查询,在系统后端返回的实体候选列表中选择实体;

(3) 系统前端将用户选择的实体记录并显示,作为关系搜索的端点.用户可继续重复步骤(2)到步骤(3),从而选择多个实体作为搜索端点;

(4) 在所选择的搜索端点大于等于两个时,用户可

以开始实体关系搜索;

- (5) 系统前端将所选择的搜索端点集合发回后端;
- (6) 系统后端根据数据关联网络图模型, 计算实体搜索端点集合之间的关联网络图;
- (7) 系统后端将结果数据发给前端;
- (8) 系统前端根据后端返回数据生成关联网络图, 关联网络图中包含搜索端点; 若某两个搜索端点之间不存在关系, 则告知用户;
- (9) 用户继续操作该关联网络图中的任意元素, 包括搜索端点, 其操作方式和第 2.1 节相同.

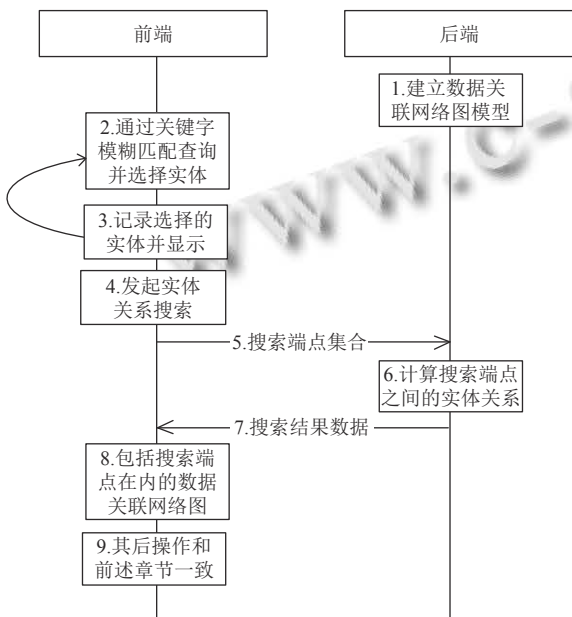


图3 搜索的前后端交互流程

步骤(6)采用分布式计算扩展关联节点并计算最小连通图算法, 其主线程与 WorkAgent 及 WorkQueue 关系如图 4, 其主进程具体步骤如图 5.

(1) 系统主进程将实体节点插入 WorkQueue 中, 在搜索开始时, WorkQueue 实体节点只有前端选择的搜索端点集合; WorkQueue 是一个工作队列.

(2) 系统主进程根据系统计算性能, 开启多个 WorkAgent, 每一个 WorkAgent 都是独立工作的, 可以采用多线程或分布式计算技术.

(3) 主进程启动定时器.

定时流程完成如下工作, 如图 6 所示.

- (1) 取 WorkQueue 中的全部节点;
- (2) 包括所有搜索端点在内的最小连通图判断, 如果不能构成最小连通图, 则执行步骤(3), 如果可以构

成最小连通图, 则执行步骤(6);

(3) 判断是否满足可配置设定的停止条件, 如果不满足停止条件, 则执行步骤(4), 如果满足, 则执行步骤(6);

(4) 将本次计算的节点插入已经构建的关系图并保存;

(5) 设置并等待下一次定时计算;

(6) 对所有 WorkAgent 发送停止信号, 搜索结束.

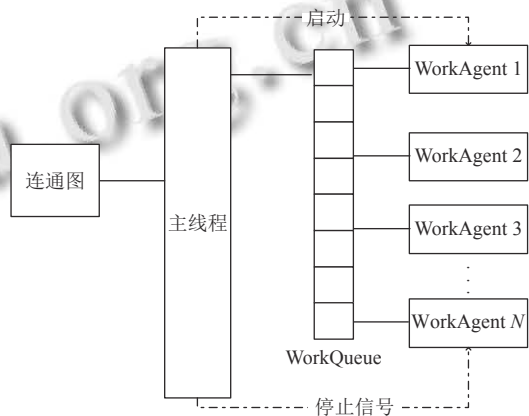


图4 主线程/WorkAgent

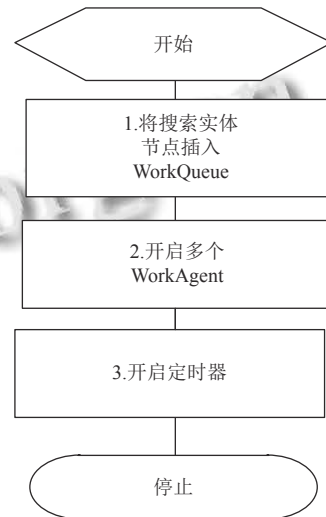


图5 主进程流程

在定时流程工作步骤(3)中所述的停止条件, 可以配置, 停止条件可以包括:

- (1) 节点总数, 超过设定的查询节点总数即停止;
- (2) 查询深度, 超过设定的从搜索端点开始的查询深度即停止;
- (3) 查询时间, 超过设定的查询时间即停止.

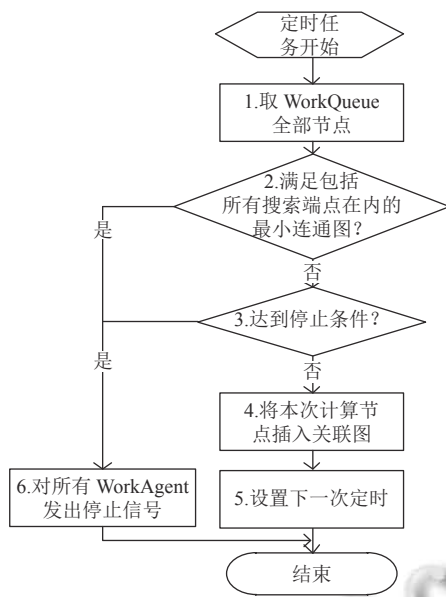


图6 定时流程

主线程启动的 WorkAgent 完成如下工作, 如图 7. 其中, 虚线表示步骤 5 是其他线程 (主线程) 发给本线程 (WorkAgent 线程) 的信号, 实线表示本线程中的步骤.

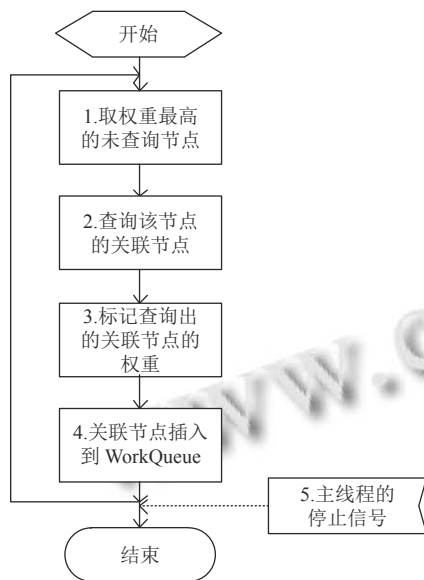


图7 WorkAgent 流程

- (1) 根据权重从 WorkQueue 中选择一个权重最高的未查询过节点;
- (2) 查询该节点的关联节点;
- (3) 标记查出的关联节点权重;
- (4) 将这些节点插入到 WorkQueue.

重复步骤 (1) 到步骤 (2), 直到接收到主进程的停止信号. 在 WorkAgent 工作步骤中所述的权重, 可以静态配置或者动态计算. 如果以静态配置方式, 则可以根据不同的节点间关系类型设定权重大小.

3 动态数据关联网络的实例和应用实验

本数据关联网络表示和搜索系统中使用的实体、关系和实体聚合可以根据不同的应用场景实例化为不同的实际对象, 其中较能体现本方法和系统的是应用在从单纯的人与人的社会网络拓展出的, 包括人、物、企事业单位等多种实体、较为复杂的数据关联网络.

3.1 动态数据关联网络的表示实例

数据关联网络的实体元素包括人员、汽车、房产、学校等; 关系包括配偶双向关系、父亲/母亲单向关系、车主单向关系、房主单向关系、工作单位单向关系等; 实体聚合包括孩子聚合、房产聚合、车产聚合等. 具体用户操作, 其在前端上的体现可以如下所述.

用户根据关键字模糊匹配搜索某种类型的实体, 如对“人员”类实体进行搜索, 关键字为“张”. 也可针对企业、房屋、车辆等其他任何种类实体进行搜索. 系统后端返回模糊匹配后的“人员”实体列表, 如图 8 所示.

选择所关心的实体作为入口, 例如“张国强”, 则系统显示和该实体直接关联的实体及其关系, 如果同类实体较多, 则显示实体聚合, 如图 9 所示. “张国强”作为当前聚焦实体, 在右边的属性/列表栏显示和其相关的属性, 包括简单属性, 如姓名、性别等, 也可包括列表属性, 如其迁徙记录等. 在“张国强”实体周围, 显示和它直接相关的实体, 并用有向线表示实体之间的关系, 如与“韩昌凤”是双向的“配偶”关系, 与“巴**童装厂”是单向的“工作单位”关系等. 如果同类实体较多, 则显示实体聚合, 比如有两套房, 则显示房产聚合; 有两个孩子, 则也显示聚合; 实体聚合上的数字表示实体个数, 例如车产实体聚合上的 8 表示“张国强”拥有 8 辆车.

点击当前显示的任何一个实体, 则显示该实体属性, 并扩展出该实体直接关联的实体、关系和实体聚合. 比如点击“韩昌凤”实体, 则当前聚焦实体变为“韩昌凤”, 在右边的属性/列表栏显示和其相关的属性, 关系网络图扩展显示“韩昌凤”的直接关联的实体和实体聚合, 如图 10.

#	姓名	性别	年龄	身份证号
1	张国强	男	35	3305*****303
2	张青涛	男	55	3305*****202
3	张广龙	男	10	3305*****581
4	张婷婷	女	7	3305*****201
5	张尔	女	45	3305*****106
6	张恒	女	45	3305*****541
7	张宁	女	30	3305*****202
8	张牛	女	32	3305*****192
9	张悦浩	女	31	3305*****921
10	张小明	男	35	3305*****303

总计 13 个

图 8 模糊搜索结果示意

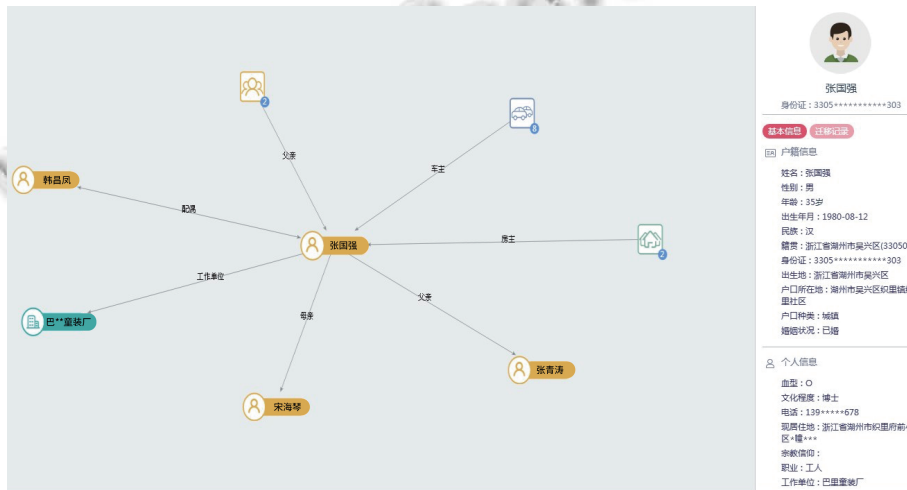


图 9 关联网络示意图 1 (点击“张国强”)

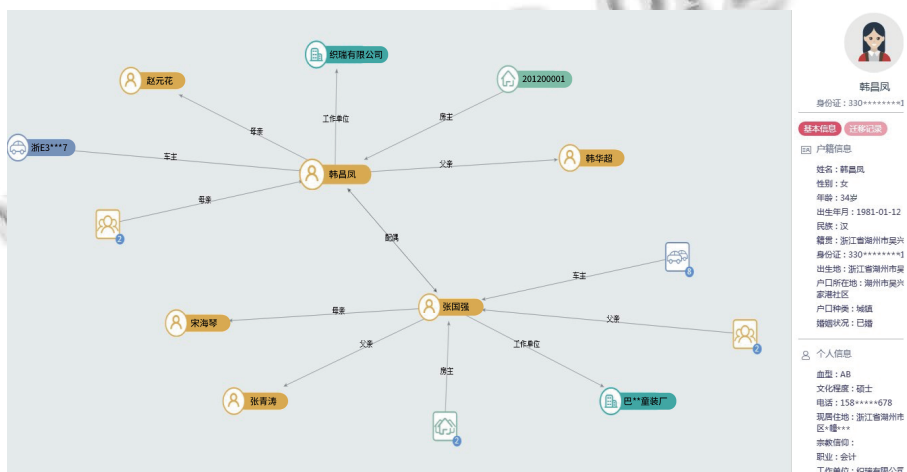


图 10 关联网络示意图 2 (点击“韩昌凤”)

点击当前显示的任何一个实体聚合, 则显示该实体的列表. 例如点击“张国强”所属的车辆实体聚

合, 在右边属性/列表栏显示所有该实体聚合的车辆列表, 如图 11 所示.

选择该列表中的实体,则显示实体,如选择车辆“浙E2***6”,则显示该车辆,并在右边属性/列表栏显

示该车辆的相关属性,包括简单属性和列表属性,如违章记录等。如图12所示。

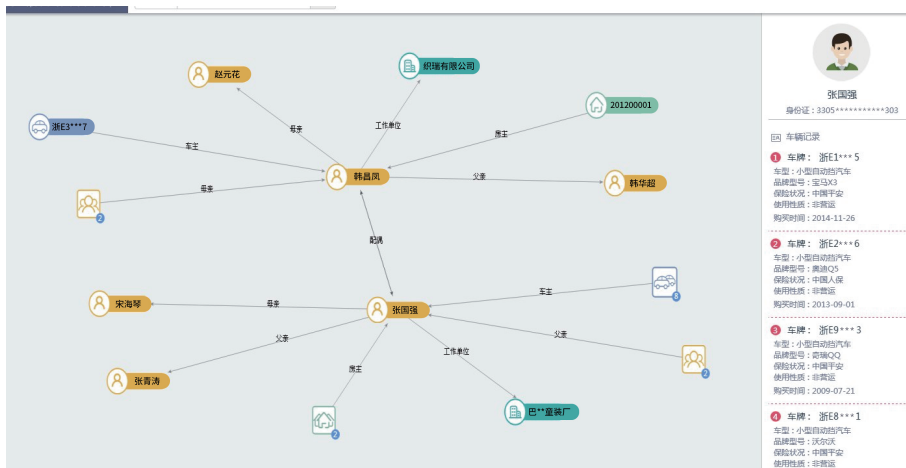


图 11 关联网络示意图 3 (点击车辆实体聚合)

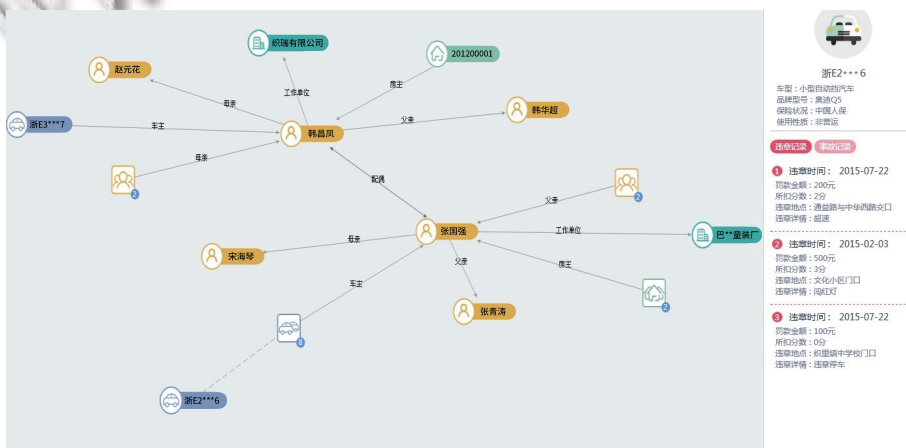


图 12 关联网络示意图 4 (点击“浙E2***6”)

继续类似步骤,点击不同实体或者实体聚合,则关联网络图可以不断地扩展,从而动态扩展表示不同实体之间的关联关系,并构成关联网络图,一种状态如图13所示。理论上,通过用户的不断交互式,该关联网络图可以无限地扩展。

3.2 数据关联网络搜索的应用实例

在更多情况下,用户已知几个不同的实体,可以是人、物、企事业单位等,需要发掘出这些人或物的直接或者间接的关联。用户根据关键字模糊匹配搜索某种类型的实体,重复操作,选择两个或以上,作为搜索端点。在本例中先对“人员”类实体进行以“张”为关键字的搜索,并在结果列表内选择“张国强”,此时前端显

示“张国强”作为搜索端点;类似的,对“企业”类实体进行搜索,并选择“巴**童装厂”;对“车辆”类实体进行搜索,并选择“浙E7***Y”。此时共有3个搜索端点,如图14。

点击搜索按钮,开始实体关系搜索,前端显示包括3个搜索端点在内的数据关联网络图,其中图标内带五角星的即搜索端点,如图15所示。从图中可见,“张国强”和“巴**童装厂”关联是直接的,前者的工作单位为后者;而“浙E7***Y”与其他两个搜索端点的关系是间接的,它是“巴**童装厂”的法人“李燕峰”的配偶“朱月娥”的车产。本搜索方法把3个搜索端点之间直接或者间接的关系直观的通过数据关联网络图实现出来。

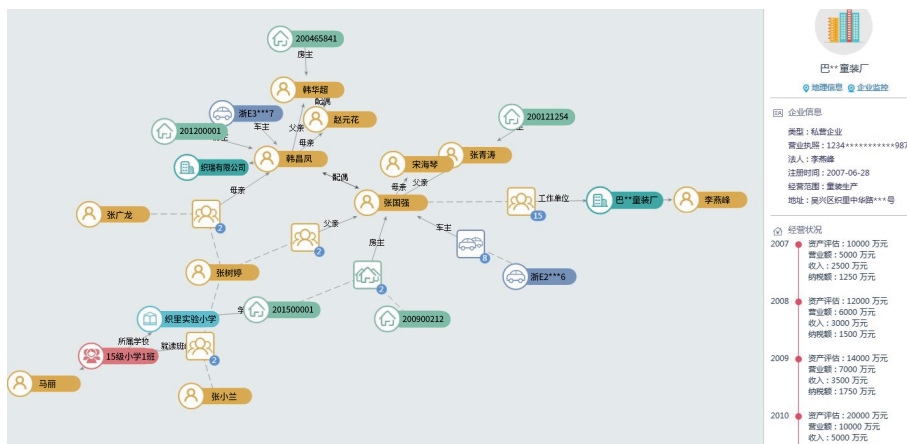


图13 关联网络示意图5 (不断交互式扩展)

#	车牌号	车辆型号
1	浙E1***5	宝马X3
2	浙E2***6	奥迪Q5
3	浙E3***7	MINI COOPER
4	浙E3***5	奔驰S500
5	浙E3***7	宝马740
6	浙E8***C	宝马760
7	浙E7***Y	奥迪Q7
8	浙E8***Y	宝马X6
9	浙E9***5	宝马Z4
10	浙E6***9	布加迪威龙

总计 169 个

图14 人、企、车3个(类)搜索端点

点击当前任何一个元素,则操作和前述数据关联网络表示方法一致.图16是点击“朱月娥”实体的情形,

图17是继续点击“李俊”实体的情形.此关联网络图不断扩展,其中3个搜索端点“张国强”“巴**童装厂”和“浙E7***Y”的图标一直带五角星符号.

3.3 应用实验

在实验室对本方法和系统进行实验测试.表1是实验系统的配置情况.本系统采用Java语言编写,数据库采用了MySQL,每个实体类型被设计为一张单表.通过模拟程序生成测试数据,测试数据被模拟为不同的实体类型,并根据模拟程序随机生成不同实体之间的关联关系.当数据总量增加时,每个不同实体类型的数量同比例增加,其关联关系也同比例增加.实体类型即模拟上述的人员、汽车、房产、学校等,关联关系即模拟上述的配偶双向关系、父亲/母亲单向关系、车单向关系、房主单向关系、工作单位单向关系等.

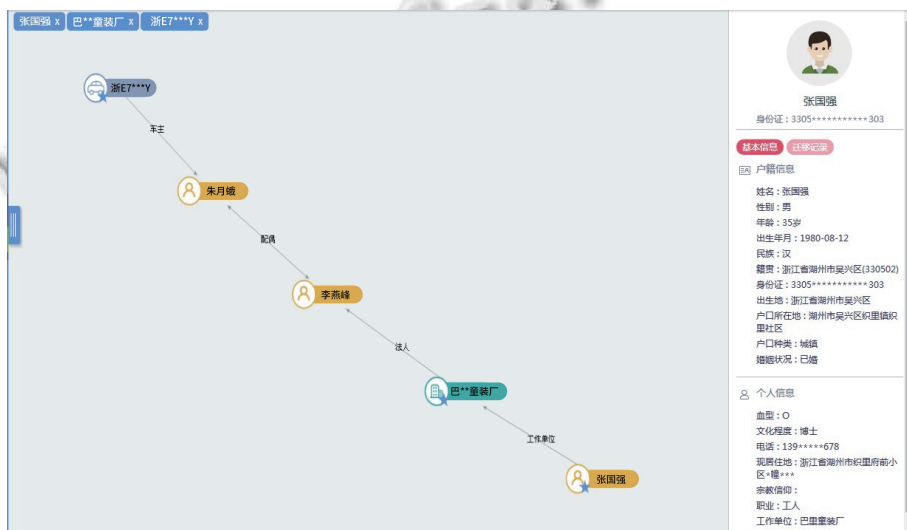


图15 搜索端点结果的数据关联展示

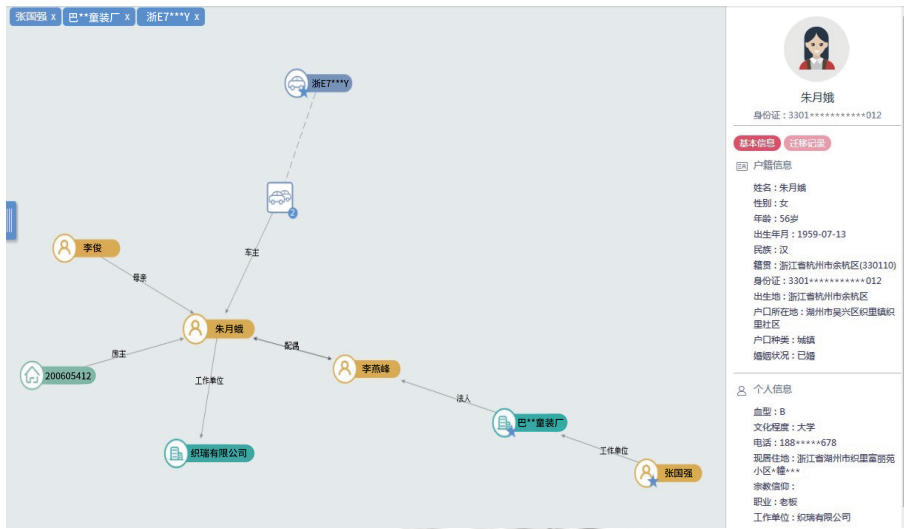


图 16 数据关联网络扩展 1 (点击“朱月娥”)

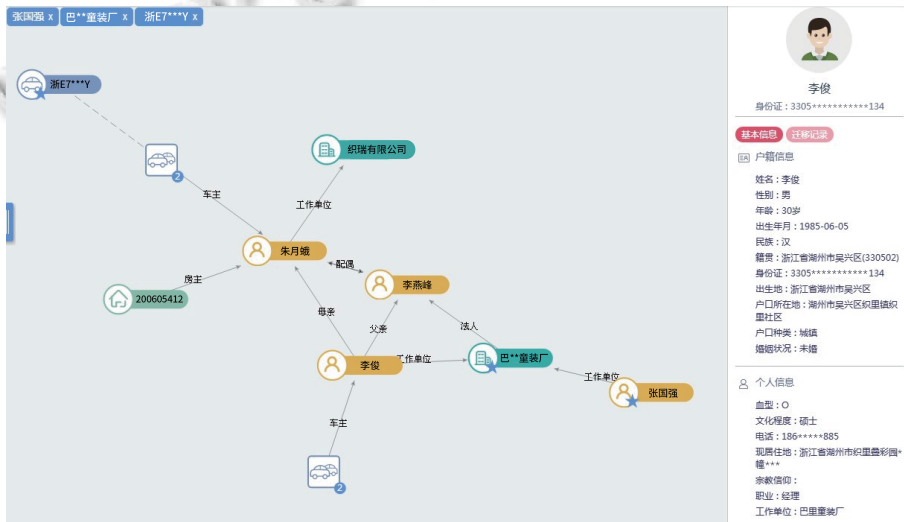


图 17 数据关联网络扩展 2 (点击“李俊”)

表 1 实验环境配置

类别	型号/版本
数据库环境	CPU 8×Intel(R) Xeon(R) Gold 6230R @ 2.10 GHz
	内存 16 GB
	操作系统 Ubuntu 18.04
	数据库 MySQL 8.0.25
主程序环境	CPU 4×Intel(R) Xeon(R) E5-2630 v3 @ 2.40 GHz
	内存 8 GB
	操作系统 CentOS 7.2
	JVM环境 1.8.0_171

表 2 是实验中数据关联网络如第 3.1 节所述扩展表示的每次单步扩展的平均时间. 从该表可看出, 随着数据总量增加, 单步扩展平均时间也有所增加, 但变化不算大. 由于每次单步的扩展实际上是一次关系式数据

表查询, 在建立适当索引的情况下, 数据总量的增加, 即每个单表同比例增加, 在实验的数据范围内对查询影响不太大. 而实体类型数量对单步扩展平均时间影响不明显.

表 3 是实验中数据关联网络如第 3.2 节所述关联搜索的平均时间. 从测试结果可以看出, 随着数据总量增加, 搜索的平均时间有明显的增加, 但与关系类型数量增加没有明显关系. 增大前述 WorkAgent 的数量, 则会对搜索速度有明显改善.

3.4 在实际工程中的应用

本数据关联网络系统中使用的实体、关系和实体聚合可以根据不同的应用场景实例化为不同的实际对象, 因此可以广泛地适应不同应用场景的要求. 基于本方法和系统开发的具体应用已经被使用到很多落地工程项目中.

例如在某智慧/平安城市项目中,使用本系统可如上面数据关联的应用实例所描述,帮助管理部门生动形象、可互动的获得其关心的人、物、单位的周边关系,并且在已知几个可能有关的人、物、单位时,通过数据关联网络的搜索得到其实际直接或者间接关系。

再如,在某智慧城市管理项目中,以及在某城域物联专网项目中,其实体或实体聚合可以是城市管理中的管理人员、城市部件、关联单位等,通过类似的手段,可以很好帮助用户提高管理效率,提升管理体验。

表2 数据关联网络表示单步扩展平均时间

实体数据总量(条)	实体类型数量(类)	单扩展平均时间(ms)
100 000	5	234
500 000	5	278
1 000 000	5	299
5 000 000	5	410
100 000	10	245
500 000	10	256
1 000 000	10	320
5 000 000	10	421

表3 数据关联网络表示查询平均时间

实体数据总量(条)	实体类型数量(类)	WorkAgent数量(个)	搜索平均时间(ms)
500 000	5	6	1472
1 500 000	5	6	3 217
5 000 000	5	6	9 084
500 000	10	6	1 432
1 500 000	10	6	3 373
5 000 000	10	6	9 423
500 000	5	12	947
1 500 000	5	12	1 955
5 000 000	5	12	6 472
500 000	10	12	983
1 500 000	10	12	2 160
5 000 000	10	12	6 691

4 结论与展望

本文提出和实现了一种动态引导式交互的数据关联网络表示和搜索方法和系统,能够在数据实体较多、关联关系较复杂时,帮助用户通过引导式交互、生动形象地获得其关心的实体的周边关系,并且在已知几个实体时候,发掘出它们之间可能存在的直接或者间接的关系。本文所描述的方法和系统已经在多个实际项目中获得使用,取得了较好的实际效果。

参考文献

- 曾悠. 大数据时代背景下的数据可视化概念研究 [硕士学位论文]. 杭州: 浙江大学, 2014.
- 孙秋年, 饶元. 基于关联分析的网络数据可视化技术研究

综述. 计算机科学, 2015, 42(S1): 484-488.

- Bizer C, Heath T, Berners-Lee T. Linked data—The story so far. *International Journal on Semantic Web Information Systems*, 2009, 5(3): 1-22. [doi: 10.4018/jswis.2009081901]
- Yue P, Guo X, Zhang MD, et al. Linked data and SDI: The case on web geoprocessing workflows. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 114: 245-257. [doi: 10.1016/j.isprsjprs.2015.11.009]
- Lee BT. Design issues: Linked data. <http://www.w3.org/DesignIssues/LinkedData.html>. (2018-06-06).
- 陈为, 沈则潜, 陶煜波. 数据可视化. 北京: 电子工业出版社, 2013.
- Freeman LC. Visualizing social networks. *Journal of Social Structure*, 2000, 1(1): 1-10.
- Yi JS, Ah Kang Y, Stasko JT, et al. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2007, 13(6): 1224-1231. [doi: 10.1109/TVCG.2007.70515]
- Liu ZC, Stasko JT. Mental models, visual reasoning and interaction in information visualization: A top-down perspective. *IEEE Transactions on Visualization and Computer Graphics*, 2010, 16(6): 999-1008. [doi: 10.1109/TVCG.2010.177]
- Keim D A, Kohlhammer J, Ellis G, et al. Mastering the Information Age—Solving Problems with Visual Analytics. Goslar: Eurographics Association, 2010.
- 丁楠, 王钰, 潘有能. 基于关联数据的政府信息聚合研究. *情报理论与实践*, 2015, 38(7): 76-79, 85.
- 徐长志, 姜良存, 乐鹏, 等. 基于关联数据的政务数据共享与空间关联研究. *测绘地理信息*, 2020, 45(6): 145-149.
- 王柳, 汤庸, 杨佐希, 等. 基于学者社交网络的论文与项目关联模型. *计算机应用研究*, 2020, 37(5): 1428-1431.
- 王晰巍, 张柳, 韦雅楠, 等. 社交网络舆情中意见领袖主题图谱构建及关系路径研究——基于网络谣言话题的分析. *情报资料工作*, 2020, 41(2): 47-55.
- 杜晓林. 大规模社会网络可视化若干问题及算法研究 [博士学位论文]. 哈尔滨: 哈尔滨工业大学, 2015.
- Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Physical Review E*, 2004, 69(2): 026113. [doi: 10.1103/PhysRevE.69.026113]
- Travers J, Milgram S. An experimental study of the small world problem. *Sociometry*, 1969, 32(4): 425-443. [doi: 10.2307/2786545]
- Newman MEJ, Strogatz SH, Watts DJ. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 2001, 64(2): 026118. [doi: 10.1103/PhysRevE.64.026118]

(校对责编: 孙君艳)