

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2017 年 7 月 6 日 (06.07.2017)



(10) 国际公布号
WO 2017/114455 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2016/112957
- (22) 国际申请日: 2016 年 12 月 29 日 (29.12.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201511031504.0 2015 年 12 月 31 日 (31.12.2015) CN
- (71) 申请人: 中国银联股份有限公司 (CHINA UNION-PAY CO., LTD.) [CN/CN]; 中国上海市浦东新区含笑路 36 号银联大厦, Shanghai 200135 (CN)。
- (72) 发明人: 何东杰 (HE, Dongjie); 中国上海市浦东新区含笑路 36 号银联大厦 7 楼, Shanghai 200135 (CN)。
- (74) 代理人: 中国专利代理(香港)有限公司 (CHINA PATENT AGENT (HK) LTD.); 中国香港特别行政区香港湾仔港湾道 23 号鹰君中心 22 字楼, Hong Kong (CN)。

- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

(54) Title: DATA PROCESSING METHOD AND SYSTEM BASED ON GRAPH

(54) 发明名称: 一种基于图的数据处理方法和系统

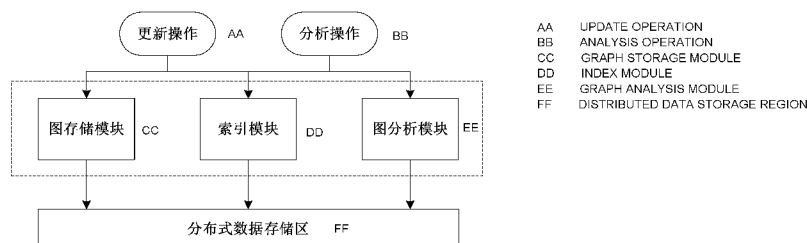


图 1

(57) Abstract: A data processing method based on a graph, for use in maintaining a graph that uses objects as vertexes and uses correlative information among objects as sides. The method comprises: maintaining vertex information and side information, the vertex information being key information about objects; and maintaining an index that uses one item of the key information of each object as a key and uses position information of the object in a graph as auxiliary information.

(57) 摘要: 一种基于图的数据处理方法, 维护以对象为顶点, 对象与对象之间的关联信息为边的图, 该方法包括: 维护顶点信息和边信息, 所述顶点信息是关于对象的关键信息, 维护以对象的关键信息中的一项为键、以对象在该图中的位置信息为辅助信息的索引。

WO 2017/114455 A1

一种基于图的数据处理方法和系统

技术领域

本发明的实施例涉及数据分析，并且具体地涉及基于图的数据处理方法和系统。

背景技术

现有的数据分析或者数据挖掘方法通常采用离线分析的方式。传统的数据处理模型和算法的缺点包括：不能在获取实时数据的同时，对数据进行有效的分析；以及不能快速地响应外部请求，基于最新的数据对外提供服务。

例如，现有的方案在对外提供数据服务时，往往需要对数据进行离线预处理和加工，将分析的结果存储在关系型数据库中。然后，响应于外部的请求，从该关系数据库中读取分析的结果。数据库中的分析结果需要被定期更新。在大数据场景下，这种方案会降低数据分析服务的效率和时效性。

发明内容

根据本发明的一个实施例，公开一种基于图的数据处理方法，维护以对象为顶点，对象与对象之间的关联信息为边的图，该方法包括：维护顶点信息和边信息，所述顶点信息是关于对象的关键信息，维护以对象的关键信息中的一项为键、以对象在该图中的位置信息为辅助信息的索引。

根据本发明的一个实施例，当对象的关键信息发生变化时，实时地更新该对象在该图中的顶点信息。

根据本发明的一个实施例，通过第一对象的关键信息利用该索引定位该第一对象在该图中的位置，根据该第一对象在该图中的位置找出与该第一对象关联的其它对象。

根据本发明的一个实施例，通过分布式架构存储所述图和所述索引。

根据本发明的一个实施例，所述对象包括银行卡用户和商户。

根据本发明的一个实施例，公开一种基于图的数据处理系统，该系统被配置成维护以对象为顶点，对象与对象之间的关联信息为边的图，该系统包括：图存储模块，用于维护顶点信息和边信息，该顶点信息是对象的关键信息，索引模块，用于维护以对象的关键信息中的一项为键、以对象在该图中的位置信息为辅助信息的索引。

根据本发明的一个实施例，该系统包括：数据更新模块，其被配置成当对象的关键信息发生变化时，实时地更新该对象在该图中的顶点信息。

根据本发明的一个实施例，该系统包括：图分析模块，其被配置成：通过第一对象的关键信息利用该索引模块定位该第一对象在该图中的位置，基于该图存储模块，根据该第一对象在该图中的位置找出与该第一对象关联的其它对象。

根据本发明的一个实施例，该系统被配置成：通过分布式架构存储所述图和所述索引。

本发明的技术方案通过缩短数据更新以及数据分析的时间，有效提升数据的时效性。有效解决了传统技术在大数据分析和数据服务方面的时效性问题。在大数据服务的场景下，更大地发挥数据的价值。

当结合附图阅读以下描述时也将理解本发明的实施例的其它特征和优势，其中附图借助于实例示出了本发明的实施例的原理。

附图说明

图 1 是根据本发明实施例的基于图的数据处理系统的示意图。

图 2 是根据本发明实施例的基于图的数据处理方法的流程图。

具体实施方式

在下文中，将结合实施例描述本发明的原理。应当理解的是，给出的实施例只是为了本领域技术人员更好地理解并且实践本发明，而不是限制本发明的范

围。例如，本说明书中包含许多具体的实施细节不应被解释为对发明的范围或可能被要求保护的范围的限制，而是应该被视为特定于实施例的描述。例如，在各实施例的上下文描述的特征可被组合在单一实施例中实施。在单一实施例的上下文中描述的特征可在多个实施例来实施。

本发明提出基于图存储模型对要处理的数据进行实时存储和更新。图是一种数据结构，定义为： $\text{graph} = (V, E)$ 。V 是一个非空有限集合，代表顶点（节点），E 代表边的集合，一般用 (V_x, V_y) 表示，其中， V_x, V_y 属于 V。若两个结点 U、V 之间有一条边连接，则称这两个结点 U、V 是关联的。可以用带权图表示两个相邻顶点之间的除连接关系以外的其它关系。基于这样的概念，本发明提出维护以对象为顶点，对象与对象之间的关联信息为边的图，以便基于图的关联模型和算法进行对象（个体）之间的关联分析来提高数据分析的性能和效率。在本发明中，个体可以是例如银行卡的用户或者任何使用网络服务（例如，网上购物）的用户。

图 1 是根据本发明实施例的基于图的数据处理系统的示意图。如图 1 所示，以虚线示出的基于图的数据处理系统，其可以包括图存储模块、索引模块、图分析模块。

图存储模块，用于维护顶点信息和边信息，该顶点信息是对象的关键信息。例如，当对象是银行卡用户时，对象的关键信息可以包括用户 ID、用户年龄范围、或者消费事件；例如，当对象是商户时，对象的关键信息可以包括商户 ID、商户类型等。边信息对象与对象之间的关联信息，例如基于用户和商户之间的消费事件产生的关联。如图所示，可以基于分布式架构存储所述图的顶点和边信息。

索引模块，用于维护以对象的关键信息的一项（例如，用户 ID 或者商户 ID）为键、以对象在该图中的位置信息为辅助信息的索引。这里，位置信息指示该对象所对应的顶点在图的存储结构（例如，邻接矩阵、邻接表等）中的与其它对象的位置关系。图分析模块通过索引能够快速定位对象在图中的位置。如图所示，可以基于分布式架构存储所述索引。

图分析模块，其被配置成通过第一对象的关键信息利用该索引模块定位该第

一对象在该图中的位置；以及基于该图存储模块，根据该第一对象在该图中的位置找出与该第一对象关联的其它对象。优选地，可以利用与图相关的各种算法来分析该对象。

基于图存储模块、索引模块、图分析模块可以高效地进行更新操作和分析操作。在更新操作过程中，当对象的关键信息发生变化时，实时地更新该对象在该图中的顶点信息。可选地，可以在基于图的数据处理系统设立更新模块。在分析操作过程中，可以接收外部请求，然后根据该请求包含的用户 ID，通过索引模块得到该用户在图中的位置信息，然后利用该位置信息借助图分析模块的计算得到分析结果。得到的分析结果作为服务信息对外部提供。

示例性实施例可在硬件、软件或其组合中来实施。例如，本发明的某些方面可在硬件中实施，而其它方面则可在软件中实施。尽管本发明的示例性实施例的方面可被示出和描述为框图、流程图，但很好理解的是，这里描述的这些装置、或方法可在作为非限制性实例的系统被实现为功能模块。此外，上述装置不应被理解为要求所有的实施例中进行这种分离，而应该被理解为所描述的程序组件和系统通常可以被集成在单一的软件产品中或打包成多个软件产品。

图 2 是根据本发明实施例的基于图的数据处理方法的流程图。如图所示，在步骤 201 中，维护以对象为顶点，对象与对象之间的关联信息为边的图。在步骤 202 中，根据外部请求中的顶点信息，定位对象在图中的位置。在步骤 203 中，根据定位的位置，分析该对象与其它对象之间的关系。

优选地，在步骤 201 中，维护顶点信息和边信息，所述顶点信息是关于对象的关键信息，以及维护以对象的关键信息中的一项为键、以对象在该图中的位置信息为辅助信息的索引。

优选地，在步骤 202 中，当对象的关键信息发生变化时，实时地更新该对象在该图中的顶点信息。

优选地，在步骤 203 中，通过第一对象的关键信息利用该索引定位该第一对象在该图中的位置，根据该第一对象在该图中的位置找出与该第一对象关联的其

它对象。例如，一个用户对象可以与多个商户对象关联。

图 2 所示的各个框可被视为方法步骤、和/或被视为由于运行计算机程序代码而导致的操作、和/或被视为构建为实施相关功能的多个耦合的逻辑电路元件。尽管操作按特定的顺序在图中被描绘，但这不应被理解为要求按照所示的特定顺序或按依次顺序来执行这些操作，或要求所有例示的操作被执行，以达到理想的结果。在某些情况下，多任务并行处理可能是有利的。

以下是基于图的数据处理方法和系统进行数据分析的示例过程。

A. 数据实时写入更新

1. 通过对外提供的数据写入接口服务将数据写入系统，该接口服务通过分布式架构对外提供高并发的数据写入能力；

2. 以写入数据的对象的关键信息为图存储中顶点的信息，通过图存储模块进行图的更新，然后将数据写入到分布式存储器中；

3. 以写入数据的对象的关键信息的一项为索引关键字，以该对象在图中存储的位置信息作为辅助信息，通过索引模块进行数据的写入和更新，然后将数据存储在分布式存储器中；

由此，通过索引服务构建快速的对象访问能力，通过图服务构建对象之间的关联关系，以支撑大规模数据的实时分析。

B. 数据实时读取分析

1. 通过对外提供的数据读取分析接口服务提供面向对象的实时分析服务，该接口服务通过分布式架构对外提供高并发的数据分析能力；

2. 以需要访问的对象关键信息的一项为关键字，利用索引服务快速定位该对象在图中的位置信息；

3. 通过获取的对象在图中的位置信息，对该对象进行数据分析和挖掘，例如，通过广度优先遍历算法进行对象的关联分析，快速识别该对象的潜在喜好对象（例如，合适于该用户的潜在商户）；

由此，能够通过接口服务返回外部所需要的分析结果，基于分布式架构和高

并发处理能力保障可靠的服务能力。

相关领域的技术人员当结合附图阅读前述说明书时，对本发明的前述示例性实施例的各种修改和变形对于相关领域的技术人员会变得明显。因此，本发明的实施例不限于所公开的特定实施例，并且变形例和其它实施例意在涵盖在所附权利要求要求的范围内。

权 利 要 求 书

1. 一种基于图的数据处理方法，其特征在于，维护以对象为顶点，对象与对象之间的关联信息为边的图，该方法包括：

维护顶点信息和边信息，所述顶点信息是关于对象的关键信息，

维护以对象的关键信息中的一项为键、以对象在该图中的位置信息为辅助信息的索引。

2. 如权利要求 1 所述的方法，其特征在于，该方法包括：

当对象的关键信息发生变化时，实时地更新该对象在该图中的顶点信息。

3. 如权利要求 1 所述的方法，其特征在于，该方法包括：

通过第一对象的关键信息利用该索引定位该第一对象在该图中的位置，

根据该第一对象在该图中的位置找出与该第一对象关联的其它对象。

4. 如权利要求 1 所述的方法，其特征在于，该方法包括：

通过分布式架构存储所述图和所述索引。

5. 如权利要求 1 所述的方法，其特征在于，

所述对象是银行卡用户。

6. 一种基于图的数据处理系统，其特征在于，该系统被配置成维护以对象为顶点，对象与对象之间的关联信息为边的图，该系统包括：

图存储模块，用于维护顶点信息和边信息，该顶点信息是对象的关键信息，

索引模块，用于维护以对象的关键信息中的一项为键、以对象在该图中的位置信息为辅助信息的索引。

7. 如权利要求 6 所述的系统，其特征在于，该系统包括：

数据更新模块，其被配置成当对象的关键信息发生变化时，实时地更新该对象在该图中的顶点信息。

8. 如权利要求 6 所述的系统，其特征在于，该系统包括：

图分析模块，其被配置成：

通过第一对象的关键信息利用该索引模块定位该第一对象在该图中的位置，

基于该图存储模块，根据该第一对象在该图中的位置找出与该第一对象关联的其它对象。

9. 如权利要求 6 所述的系统，其特征在于，该系统被配置成：

通过分布式架构存储所述图和所述索引。

10. 如权利要求 6 所述的系统，其特征在于，

所述对象是银行卡用户。

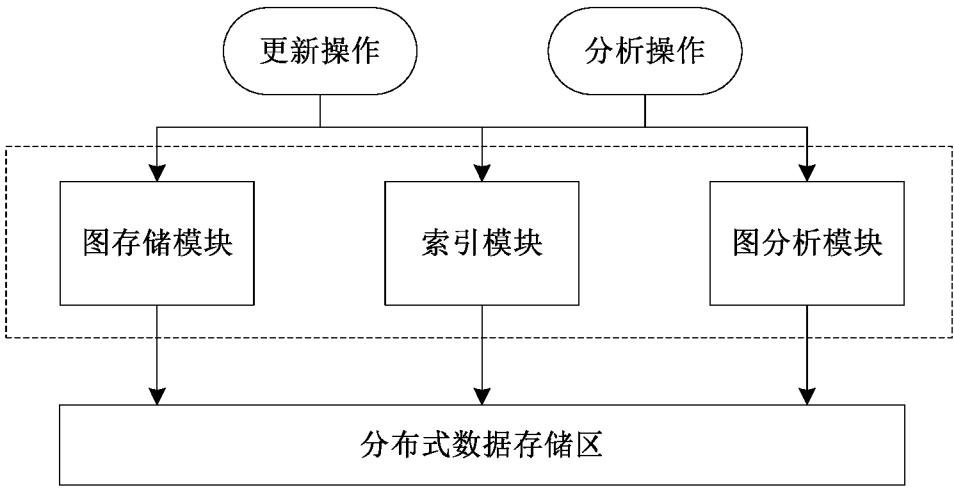


图 1

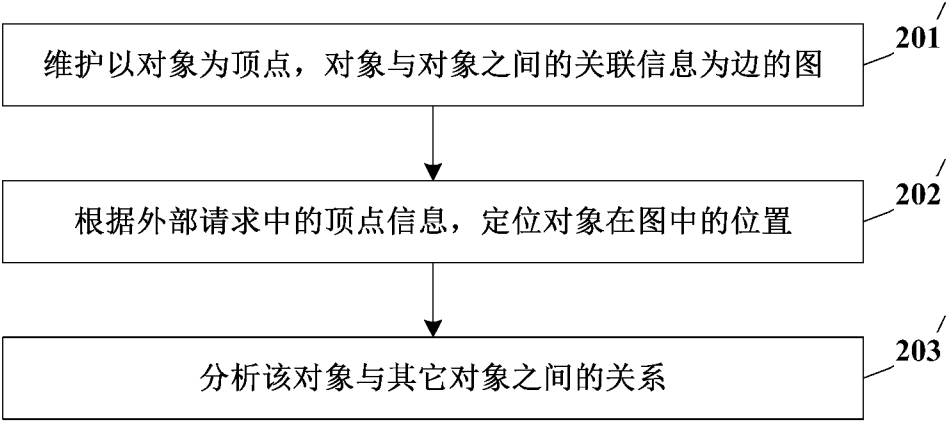


图 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/112957

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, WPI, EPODOC, CNPAT, GOOGLE; key, auxiliary information, graph, data w structure, object, vertex, edge, key w information, index, auxiliary

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 105677801 A (CHINA UNIONPAY CO., LTD.), 15 June 2016 (15.06.2016), claims 1-10	1-10
X	CN 104809249 A (BEIJING DIDI INFINITY TECHNOLOGY AND DEVELOPMENT CO., LTD.), 29 July 2015 (29.07.2015), description, paragraphs [0028]-[0029] and [0037]	1-10
A	CN 103345508 A (PEKING UNIVERSITY), 09 October 2013 (09.10.2013), the whole document	1-10
A	US 5761664 A (INTERNATIONAL BUSINESS MACHINES CORPORATION), 02 June 1998 (02.06.1998), the whole document	1-10

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 22 February 2017 (22.02.2017)	Date of mailing of the international search report 01 April 2017 (01.04.2017)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer DONG, Gang Telephone No.: (86-10) 62413649

International application No.
PCT/CN2016/112957

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2016/112957

A. 主题的分类

G06F 17/30(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNKI, WPI, EPODOC, CNPAT, GOOGLE; 图, 数据结构, 对象, 顶点, 边, 关键信息, 键, 索引, 辅助信息, graph, data w structure, object, vertex, edge, key w information, index, auxiliary

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 105677801 A (中国银联股份有限公司) 2016年 6月 15日 (2016 - 06 - 15) 权利要求1-10	1-10
X	CN 104809249 A (北京嘀嘀无限科技发展有限公司) 2015年 7月 29日 (2015 - 07 - 29) 说明书第[0028]-[0029], [0037]段	1-10
A	CN 103345508 A (北京大学) 2013年 10月 9日 (2013 - 10 - 09) 全文	1-10
A	US 5761664 A (INTERNATIONAL BUSINESS MACHINES CORPORATION) 1998年 6月 2日 (1998 - 06 - 02) 全文	1-10

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2017年 2月 22日

国际检索报告邮寄日期

2017年 4月 1日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

董刚

电话号码 (86-10)62413649

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2016/112957

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	105677801	A	2016年 6月 15日	无			
CN	104809249	A	2015年 7月 29日	无			
CN	103345508	A	2013年 10月 9日	无			
US	5761664	A	1998年 6月 2日	JP	H0756985	A	1995年 3月 3日

表 PCT/ISA/210 (同族专利附件) (2009年7月)