

面向新能源大数据的异常模式检测技术研究

吕清泉¹, 丁坤¹, 周强¹, 张健美¹, 高鹏飞¹, 张睿骁¹, 侯佳敏²

¹国网甘肃省电力公司电力科学研究院, 甘肃 兰州

²中国人民大学, 北京

Email: 376587752@qq.com

收稿日期: 2020年11月3日; 录用日期: 2020年11月18日; 发布日期: 2020年11月25日

摘要

随着电力系统规模的日益增大, 新能源的不断加入, 系统中的知识总量呈爆炸式增长, 电力系统运行需基于更高的数据质量实现, 以便为系统提供全方位, 全周期的数据共享。国内电力信息系统所使用的数据库一般为结构化数据库。而传统关系型数据库在处理大数据复杂关系问题过程中, 一系列技术瓶颈日益凸显, 传统数据库已经无法满足海量数据的处理建模与分析。本文提出了一种全自动化新能源大数据异常检测的技术方法, 它利用知识图谱天然反应数据间现有关系的优势, 基于图结构和图顶点的属性信息, 对异常图模式进行形式化定义以直接挖掘电网拓扑结构中的异常数据。本文挖掘的异常数据在现实中具有语义信息, 在异常数据检测问题上具有可行性和实用价值。通过挖掘富有语义信息的异常图模式, 检测新能源大数据中的异常数据, 以保证数据的可靠性和准确性, 避免错误或无效数据影响电力系统精细化管理和电网安全运行。算例实验效果良好, 表明所提出的辨识方法具有理论价值和实际应用价值。

关键词

数据挖掘, 频繁子图, 异常检测, 新能源大数据

Research on Abnormal Pattern Detection Technology for New Energy Big Data

Qingquan Lv¹, Kun Ding¹, Qiang Zhou¹, Jianmei Zhang¹, Pengfei Gao¹, Ruixiao Zhang¹, Jiamin Hou²

¹State Grid Gansu Electric Power Company Electric Power Research Institute, Lanzhou Gansu

²Renmin University of China, Beijing

Email: 376587752@qq.com

文章引用: 吕清泉, 丁坤, 周强, 张健美, 高鹏飞, 张睿骁, 侯佳敏. 面向新能源大数据的异常模式检测技术研究[J]. 计算机科学与应用, 2020, 10(11): 2024-2033. DOI: 10.12677/csa.2020.1011214

Abstract

With the increasing scale of the power system and the continuous addition of new energy, the total amount of knowledge in the system has exploded. The operation of the power system needs to be realized based on higher data quality in order to provide the system with all-round and full-cycle data shared. The database used by the domestic electric power information system is generally a structured database. In the process of traditional relational databases dealing with the complex relational problems of big data, a series of technical bottlenecks have become increasingly prominent, and traditional databases can no longer satisfy the processing modeling and analysis of massive data. This paper proposes a fully automated new energy big data anomaly detection technology method, which uses the advantages of the existing relationship between the natural response data of the knowledge graph, and based on the graph structure and the attribute information of the graph vertices, the abnormal graph mode is formalized to directly mine abnormal data in the grid topology. The abnormal data mined in this paper has semantic information in reality, and has feasibility and practical value in the detection of abnormal data. By mining abnormal graph patterns rich in semantic information, abnormal data in new energy big data is detected to ensure the reliability and accuracy of the data, and prevent errors or invalid data from affecting the refined management of the power system and the safe operation of the power grid. The experimental results of the calculation examples are good, indicating that the proposed identification method has theoretical value and practical application value.

Keywords

Data Mining, Frequent Subgraphs, Anomaly Detection, New Energy Big Data

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着电力系统规模的日益增大, 新能源的不断加入, 系统中的知识总量呈爆炸式增长, 电力系统运行需基于更高的数据质量实现, 以便为系统提供全方位, 全周期的数据共享。国内电力信息系统所使用的数据库一般为结构化数据库。而传统关系型数据库在处理大数据复杂关系问题过程中, 一系列技术瓶颈日益凸显, 传统数据库已经无法满足海量数据的处理建模与分析[1]。

由于新能源的多源异构性, 现有的异常数据检测方法存在难题。对于电力系统中的异常数据检测, 现有方法分为两种, 1) 离线方法: 使用硬件设备进行现场检测辨识, 成本高且效率低; 2) 在线方法: 包括通过对计量系统中的相关信息进行分析而达到异常检测的目的, 其具有成本低、实时性高的特点, 但在线方法对数据量和数据质量的要求高, 依赖采集数据, 占用通信信道。现在尚未有一种方法能够完全解决实际新能源复杂运行场景中的异常数据检测问题。与此同时, 知识图谱可以清晰地反映出数据间的现有关系, 推理挖掘出隐藏知识, 适用于新能源大数据的异常检测[2]。近年来, 知识图谱在电力系统中的应用越来越广泛, 但仍处于初期阶段, 且现有技术都是基于已有的规则信息进行挖掘的, 尚未设计一

种知识图谱数据全自动化异常数据检测技术[3] [4] [5] [6]。

本文的主要贡献如下：

1) 形式化地定义了异常数据在知识图谱中的表现形式。

2) 提出了异常度概念来衡量数据的异常程度。

3) 设计了一种面向海量新能源知识图谱数据的异常检测方法：异常图模式挖掘算法，该方法充分利用了新能源数据的多源异构性，以及新能源数据的分布模式变化多端的特点，为新能源异常数据检测提供了灵活、高效的解决方案，对于提升新能源大数据的质量具有积极作用。

2. 相关工作

文献[7]提出了一种基于知识图谱的低压配电网拓扑结构辨识方法，该方法首先构建知识图谱，随后运用知识图谱技术对低压配电网信息系统中的户变关系进行辨识。然而该方法依赖于现有的设计规范，需要有目标地在知识图谱中进行异常数据的查找，并且该方法目前只能实现研究电网友户变关系中的用电地址关系，尚未设计一种知识图谱数据全自动化异常数据检测技术。

现有技术中，知识图谱中的异常数据挖掘可以通过对图数据进行频繁子图的挖掘，找到图中正常的频繁子结构，再通过摒除正常子结构来获取异常子结构生成。然而该技术存在两方面的弊端：1) 异常子结构是通过摒除频繁子图间接生成的，这是由于频繁子图技术只挖掘正常的子图结构，不挖掘异常的子图结构，例如文献[8]提出的 FSMBUS，文献[9]提出的 Gspan 算法，文献[10]提出的 GRAMI 算法等；2) 异常子结构只涉及子图拓扑结构，未利用图顶点丰富的属性信息，而顶点属性信息能够对异常数据进行更细粒度的划分，对异常数据的检测更加敏感。

3. 异常图模式

本文在频繁子图的基础上，结合图顶点的属性信息，对异常图模式进行形式化定义以直接挖掘知识图谱中异常子结构。异常图模式分为两种：不带属性的异常图模式和带属性的异常图模式。

在介绍异常图模式前，首先介绍一些基本概念。

3.1. 基本概念

定义 1 (属性图)属性图可以表示为四元组 $G = (V, E, L, A)$ ，其中 1) V 是一组顶点的集合；对于每个 $v \in V$ 都有携带属性 id 来表明顶点的身份；2) E 是一组边的集合， $E \subseteq V \times V$ ， E 中的每一条边可表示为 $\langle v1, v2 \rangle$ ，其关联图中的顶点 $v1$ 和顶点 $v2$ 。3) L 是标记函数，它将 G 中的每个顶点和边映射到一组标签的集合， A 是一组属性的集合， V 中的每一个顶点可以有一个或多个来自于 A 的属性。对于 $v \in V$ ，令 $v.A$ 代表顶点 v 上的属性集合。

定义 2 (一般图模式)图模式可以表示为三元组 $Q[x] = (V_Q, E_Q, L_Q)$ ，其中 1) V_Q 是一组图模式顶点的集合，表示图 G 中的顶点标签；2) E_Q 是一组图模式边的集合， $E_Q \subseteq V_Q \times V_Q$ ， E_Q 中的每一条边可表示为 $\langle v1, v2 \rangle$ ，其关联图模式中的顶点 $v1$ 和顶点 $v2$ 。3) L_Q 是一组标签的集合 4) x 是变量的集合，用来简单表示图模式中不同的顶点。图模式可以看作是图中的关系模式，只是对比关系模式，图模式能表示多个实体之间的关系，且表示的关系的拓扑结构更为复杂。

定义 3 (图模式匹配)图模式匹配是指在图 G 中存在一个子图 $G' = (V', E', L', A')$ ，它与图模式 Q 是同构的。也就是说，存在从 V_Q 到 V' 的双射函数 h ，使得 1) 对于每个节点 $u \in V_Q$ ， $L'(h(u)) = L_Q(u)$ 和 2) 当且仅当 $e' = (h(u), h(u'))$ 是 G' 和 $L'(e) =$ 的边时， $e = (u, u')$ 是 Q 的边 $L_Q(e)$ 。

定义 4 (支持度)考虑一个图 G 和一个图模式 $Q[x]$ ，用 (Q, G) 表示图 G 中图模式 Q 的所有匹配。我们

将图模式 Q 的支持度定义为图模式 Q 匹配图 G 的数目，形式化表示为：

$$\text{supp}(Q, G) = |(Q, G)|$$

当一个图模式的支持度大于一定的阈值时，我们称其为频繁图模式。

3.2. 异常图模式

为描述图的异常程度，我们形式化定义异常图模式，并将其分为两种：不带属性的异常图模式和带属性的异常图模式。不带属性的图模式即为一般图模式。带属性的图模式表示在一般图模式上添加属性约束条件的集合，属性约束条件是对图顶点属性的约束。

定义 5 (带属性的图模式)带属性的图模式形式化表示为 $Q[x] + C$ ，其中 $Q[x]$ 为一般图模式， C 为顶点变量的属性值约束集。属性约束条件是对图顶点属性的约束，表达为 $x.a = v$ ，其中 $a \in A$ ，语义上表达为顶点的属性为某一属性值常量。类似一般图模式的支持度定义，将带属性的图模式的支持度用 $\text{supp}(Q + C, G)$ 表示，语义上为大图 G 上既匹配图模式 Q 又满足属性值约束 C 的子图数目。

带属性的图模式在更小的元组上进行了约束，可以放宽一般图模式的限制，从一张大图约束放宽到了元组类级别的约束。带属性的图模式是一般图模式更细粒度的划分。为直接挖掘异常图模式，本文提出异常度的概念衡量图的异常程度。

定义 6 (异常度)本文用 abn 来描述图的异常程度。其中

- 不带属性的图模式异常度：用以下公式来衡量图的异常程度，且异常度 $\text{abn}(Q, G)$ 越小， Q 的结构越不合理。其中 Q' 为频繁的图模式， Q 为频繁图模式 Q' 添加一条边扩展生成的子图结构。该公式表示为频繁图模式在扩展边生成新的子图结构时变得不频繁以至于异常。

$$\text{abn}(Q, G) = \frac{\text{supp}(Q, G)}{\text{supp}(Q', G)}$$

- 带属性的图模式异常度：用以下公式来衡量带属性的图模式的异常程度，且异常度 $\text{abn}(C + Q, G)$ 越小，带有 C 属性值约束集合条件的子图结构 Q 越不合理。 Q 为频繁图模式， $C + Q$ 表示为频繁图模式 Q 上添加属性约束条件集合 C 的约束， $C + Q$ 图模式是在 $C' + Q$ 图模式上添加 a 属性约束条件，即， C' 可以为空集， C 不为空集。

$$\text{abn}(C + Q, G) = \frac{\text{supp}(C + Q, G)}{\text{supp}(C' + Q, G)}$$

4. 异常图模式的挖掘方法

4.1. 概述

本文提出了一种异常图模式挖掘技术。图 1 是异常图模式挖掘的总体框架图。在挖掘异常图模式之前，需要对子图拓扑结构进行适当的编码，这样每个子图都有一个唯一的编码，以便更容易地检测同构，例如标准邻接矩阵 CAM 、 Gspan 的 DFSCode 。在对图进行编码后，寻找异常图模式的第一步是生成一组不带属性的候选图模式集合，例如利用 FFSM-Join 和 FFSM-Extend 算法添加扩展边生成候选图模式；第二步，计算候选集的支持度：构造图模式匹配并计算不带属性的候选图模式的支持度；第三步，1) 生成保留不带属性的频繁图模式用于下一轮不带属性的图模式候选集生成；2) 计算异常度生成不带属性的异常图模式；第四步，构造带属性的候选图模式，计算带属性的候选图模式的支持度和异常度，生成带属性的异常图模式。第五，跳转到第一步进行下一层的异常图模式生成。

其中，第一步和第二步可以是现有技术中的任一大规模图数据的频繁子图挖掘算法，异常图模式挖掘利用频繁子图技术生成的子图匹配和支持度计算结果生成不带属性/带属性的异常图模式。

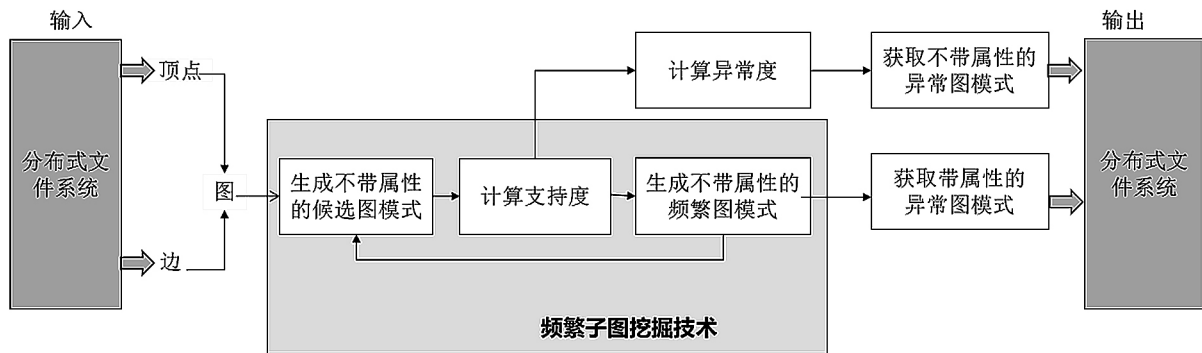


Figure 1. Overall technical route
图 1. 总技术路线

4.2. 不带属性的异常图模式

其中，一般图模式的挖掘可以是任一大规模图数据的频繁子图挖掘算法，异常图模式挖掘利用频繁子图技术生成的图模式匹配和支持度计算结果生成不带属性/带属性的异常图模式。频繁子图技术主要是在大图上对频繁的子图进行挖掘，其核心思想为：先生成候选图模式，随后对其构造图模式匹配以便进行支持度的计算，筛选出频繁的子图模式，最后再生成下一层的候选图模式，进行下一轮的频繁子图生成，直到没有新的频繁子图生成或者参数达到要求。

频繁子图的生成和具体技术本文不再赘述，不带属性的异常图模式挖掘具体算法如下所示，该算法是在频繁子图挖掘技术上的扩展，利用频繁子图中父层图模式的支持度和当前图模式的支持度，计算异常(第 5 行)，当异常度小于一定的阈值，则将其加入不带属性的异常图模式集合中(第 6 行)。算法中的第 1~3 行是带属性的异常图模式挖掘部分：若当前图模式的支持度大于支持度阈值，则认为其为频繁子图，保留其进行下一层候选子图的生成(第 2 行)，同时在频繁子图上进行带属性的异常图模式挖掘(第 3 行)，具体的带属性的异常图模式挖掘算法将在后文中进行详细描述。

算法 1: 不带属性的异常图模式挖掘

Input: support threshold m , abnormal threshold n , support, parent-support
Output: a set Σ of abnormal graph pattern
 1. if support $> m$:
 2. Merge to generate frequent graph patterns with no attributes in this layer
 3. 带属性的异常图模式挖掘(见算法 2)
 4. else
 5. if support/parent-support $< n$:/*abnormal degree*/
 6. add abnormal graph pattern to Σ

4.3. 带属性的异常图模式

为了在图模式上添加属性，我们设计以下算法流程：首先，在不带属性的频繁图模式上初始化属性约束条件生成第 $j = 1$ 层的带属性的图模式候选集(第 2 行)；第二，计算第 j 层的带属性的图模式的最小支持度(第 5 行)；第三，生成带属性的异常图模式(第 6~10 行)；最后，添加属性约束条件以生成第 $j + 1$ 层的带属性的图模式候选集(第 11 行)，跳转执行第二步直到没有新的带属性的异常图模式生成或者 j 达到阈值。其中，FAMap 为各顶点类型下的频繁属性集合；FAIDMap 为频繁属性对应的顶点 ID 集合；PFASet

为匹配图模式的频繁属性项集， nmaches 为图模式匹配。

算法 2: 带属性的异常图模式挖掘

Input: PFASet, FAIDMap, nmaches , support threshold m , abnormal threshold n
Output: a set Σ of abnormal graph pattern with attributes

1. $L_0 := \{\emptyset\}$
2. $L_1 := \{X \mid X \in \text{PFASet}\}$ /*Initialize the graph pattern candidate set of the first layer*/
3. **While** $L_j \neq \emptyset$ and $j < J$:
4. **for each** $A \in L_j$:
5. COMPUTE_SUPPORT
6. **if** support $< m$:
7. remove A from L_j /*prune*/
8. compute abnormal degree
9. **if** abnormal $< n$
10. add abnormal graph pattern with attribute A to Σ
11. $L_{j+1} := \text{GENERATE_NEXT_LEVEL}(L_j)$
12. $j := j + 1$

对于图模式第 $j + 1$ 层的属性约束条件生成，本文结合 **aprior** 算法进行属性约束条件的生成：对于第 j 层的每一个属性约束条件 A ，根据“A.顶点 v_i -属性”进行分组(第 2 行)；随后，利用组中两个不同属性条件，两两合并形成一个下一层属性约束条件(第 3~4 行)；最后，对于刚生成的 L_{j+1} 的属性约束条件进行遍历检查，确定它的属性约束条件的真子集结点是否存于 L_j 中。若存在，则将其纳入到 L_{j+1} 的属性约束集合中(第 5~6 行)。具体的算法如下：

Procedure: GENERATE_NEXT_LEVEL(L_j)

1. $L_{j+1} := \{\emptyset\}$
2. **for each** $K \in \text{SAME_BLOCK}(L_j)$
3. **for each** $\{Y, Z\} \subseteq K, Y \neq Z$
4. $X := Y \cup Z$
5. **if for all** $A \in X, X \setminus \{A\} \in L_j$
6. $L_{j+1} := L_{j+1} \cup \{X\}$
7. **return** L_{j+1}

5. 实验结果

5.1. 实验环境

实验使用的是高性能计算集群，集群操作系统为 CentOS 6，使用其中 20 个计算节点：node-0-1~16 以及 node-1-1~4，每个计算节点配置是双路刀片节点，参数是 24*Intel Xeon E5 2650 2.2 GHz/64 GB 内存。

5.2. 实验数据

在现有条件下缺乏电网的大规模数据集，而大规模知识库的构造涉及到数据融合，实体对齐等一系列复杂的数据挖掘技术，不涉及本文的核心技术点。另一方面，现有的电力数据满足不了大数据的 4v 特性。因此本文采用的是通用的大规模知识库 yago 数据集。但这并不影响对于算法在电力系统下的应用，因为图数据的格式是通用的，异常图模式的挖掘是自动化的，不涉及具体的规则依赖。

YAGO [11]是一个大型语义知识库，源自 Wikipedia, WordNet, WikiData, GeoNames 和其他数据源。

目前, YAGO 包含超过 1700 万个实体(例如个人, 组织, 城市等), 并且包含有关这些实体的 1.5 亿多个事实(元组)。YAGO3 带有 4.29 百万个顶点, 12.43 百万条边, 其中, 顶点类型有 23 类, 边类型有 37 类, 其中顶点类型是本文根据需要通过 yagoSchema (关系的定义域和值域) yagoTransitiveType (实体所有来源类型)和 yagoTaxonomy (类型子属)标注。为了探讨图的大小对算法的影响, 本文将 YAGO3 数据进行划分成 1 百万、3.5 百万、6 百万、8.5 百万以及 11 百万, 为了保证图的连续性不能随意抽取其中的数据作为最后的结果数据进行划分, 而是先随机抽取部分三元组数据, 然后再添加以这些三元组数据的主语为主语的三元组以及以这些三元组数据的宾语为主语的三元组。例如对于 1 百万的数据, 随机抽取出 10 万的三元组数据, 再添加以 10 万三元组的主语为主语的其他三元组以及 10 万三元组的宾语为主语的其他三元组来构成 1 百万的数据。

5.3. 实验结果

当图的数量 s : 1 m, 3.5 m, 6 m, 8.5 m, 11 m 时, 集中式环境下异常图模式的数量变化和运行时间如表 1 所示。实验表明当图的规模增大时, 挖掘异常图模式所需的时间越长。同时在图规模达到一定数量时, 对于异常图模式的挖掘的时间呈指数型增长, 这表明了分布式算法的必要性。

Table 1. The effect of the number of graphs

表 1. 图数量的影响

图数量	1 m	3.5 m	6 m	8 m	11 m
运行时间(s)	26.749	632.254	3988.158	12,182.461	22,073.541
异常图模式	23	83	103	143	163

如图 2 所示为集中式算法和分布式算法的时间对比, 不难发现分布式算法在大图中发现异常图函数依赖比集中式有优势, 这种优势在图越大的情况下越能够体现。特别是当图数量 = 11 百万, 运行时间集中式算法是分布式算法的 5.5 倍, 这也证实了分布式算法的必要性以及分布式算法在大图中的可行性。

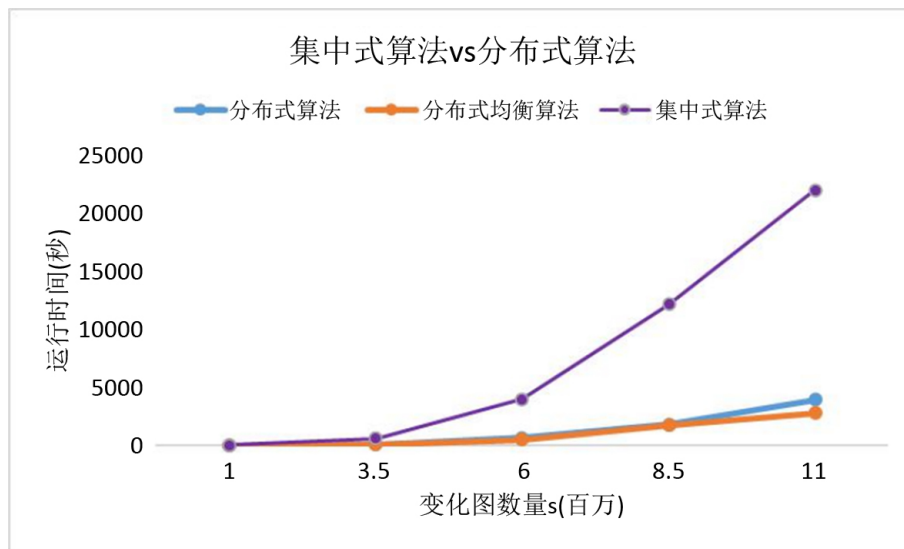


Figure 2. Time comparison curve of centralized algorithm and distributed algorithm

图 2. 集中式算法和分布式算法的时间对比曲线

变化支持度的阈值, 结合最后实验挖掘出的样例效果以及调参经验, 本文对于 yago3 异常图模式挖掘的支持度阈值设为 m : 50, 100, 150, 200, 250。其异常图模式的数量变化如表 2 所示。该实验结果

表明阈值的设定对于异常图模式的挖掘起到了较大的影响，因此在实际应用中应该考虑数据的可用性程度，适量地评估原有数据集地数据质量问题，以便挖掘出有语义的异常图模式。

Table 2. The impact of support
表 2. 支持度的影响

支持度阈值	50	100	150	200	250
运行时间(s)	3930.523	3505.003	3262.47	3106.319	2894.063
异常图模式	749	663	329	184	179

变化异常度阈值 a : 1%, 3%, 5%, 7%, 9%。异常度的选取在原则上是越小越好，基于此，本实验选择了数据集大小的 1%, 3%, 5%, 7%, 9% 大小，对其进行测评。其异常图模式的数量变化如表 3 所示。该实验结果表明异常度作为超参对于异常图模式挖掘的影响不大，这也间接表明了异常度对于异常数据的检测式具有鲁棒性的。

Table 3. The effect of anomaly
表 3. 异常度的影响

异常度阈值	1%	3%	5%	7%	9%
运行时间(s)	4051.037	3944.174	4297.114	4246.144	4435.98
异常图模式	986	992	1006	1010	1101

5.4. 案例解析

案例 1: 图 3 中的 Q_1 为一个不带属性的异常图模式，该模式表明两个人互相有孩子的图是不合理的图结构。子图 G_1 违反图模式 Q_1 在 YAGO3 的实例，Bardas_Phokas_the_Elder 和 Nikephoros_II_Phokas 互相有孩子，这不合理的，捕获了 YAGO3 中图 G_1 的错误。Nikephoros_II_Phokas 是 Bardas_Phokas_the_Elder 的孩子，但是 Bardas_Phokas_the_Elder 是 Nikephoros_II_Phokas 的父亲。

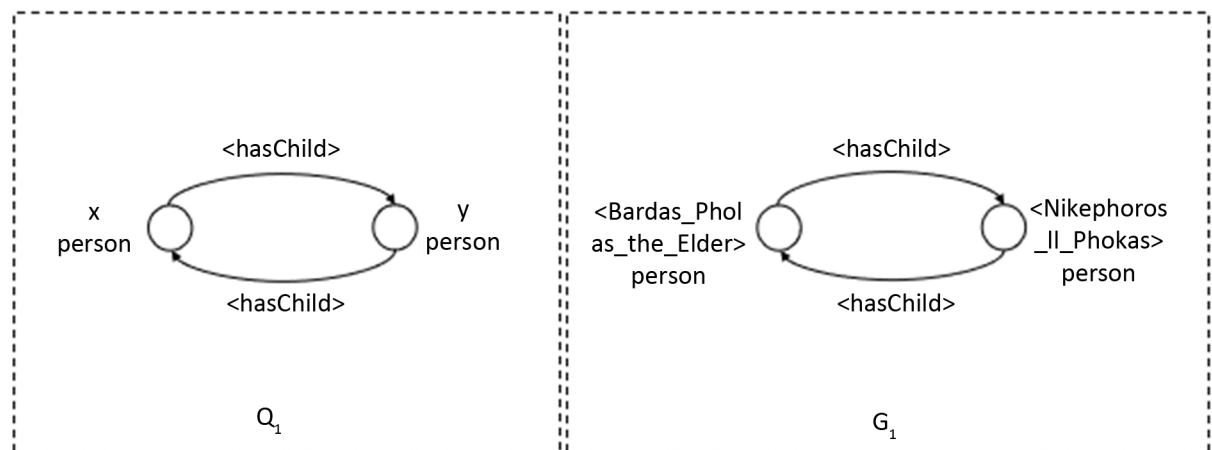


Figure 3. Exception graph pattern without attributes
图 3. 不带属性的异常图模式

案例 2: 图 4 中的 Q_2 所示为一个带属性的异常图模式，这个异常图模式表达的意思一个人既拥有德国国籍和挪威国籍的图是不合理的。子图 G_2 违反图模式 Q_2 在 YAGO3 的实例，J._A._O._Preus_II 同时拥有德国国籍和挪威国籍，这不合理的，首先挪威不支持双重国籍，德国原则上不允许双国籍身份(欧盟国

家的公民除外), 但是挪威也不是欧盟国家, 这捕也不是获了 YAGO3 中图 G_2 的错误。

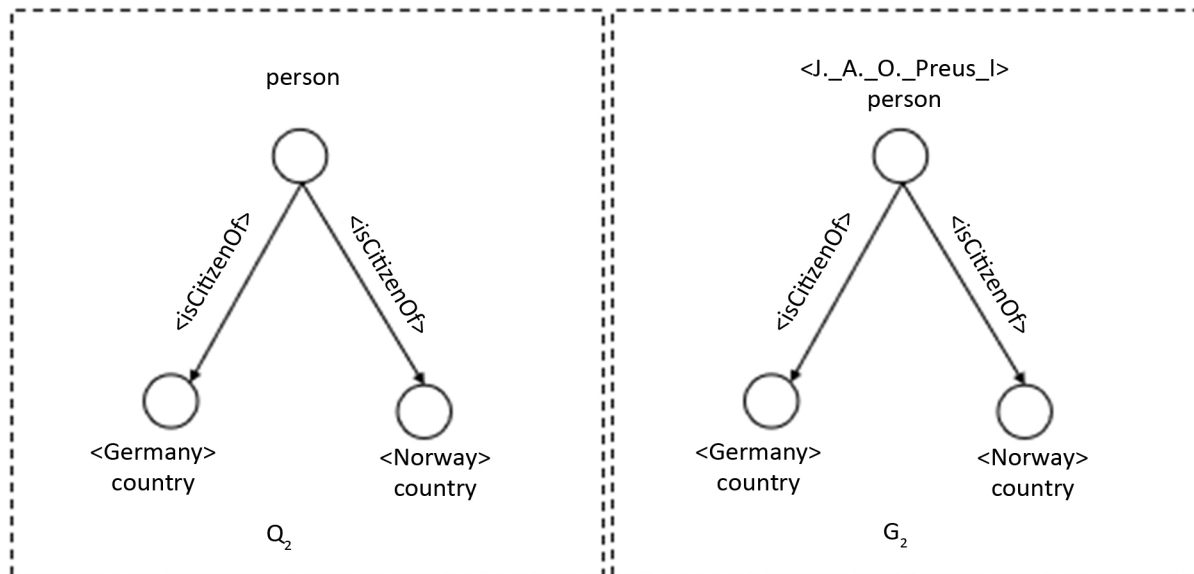


Figure 4. Exception graph pattern with attributes

图 4. 带属性的异常图模式

6. 结束语

本文基于知识图谱, 结合频繁子图挖掘技术, 利用图顶点的属性信息, 对异常图模式进行形式化定义, 并设计一种全自动化新能源大数据异常检测的技术方法, 从而自动发现知识图谱中的异常数据, 保证数据质量。在未来的工作中, 将会对新能源的真实数据进行知识图谱的构建, 并对真实数据进行更多挖掘质量的评估。

致 谢

本文工作得到国家电网有限公司科技项目(合同号: SGGSKY00FJJS1900296)的部分支持。本文工作也得到中国人民大学信息技术与管理国家级实验教学示范中心的部分支持。感谢审稿专家们的宝贵修改意见和建议, 同时感谢中国人民大学数据工程与知识工程教育部重点实验室人大行云云平台为本论文项目提供的实验环境。

参考文献

- [1] 汤亚宸, 方定江, 韩海韵, 等. 基于图数据库和知识图谱的电力设备质量综合管理系统研究[J]. 供用电, 2019, 36(11): 35-40.
- [2] 高海翔, 苗璐, 刘嘉宁, 林湘宁, 董锴, 何祥针. 知识图谱及其在电力系统中的应用研究综述[J]. 广东电力, 2020, 33(9): 66-76.
- [3] 王琼, 魏军, 闫润珍, 等. 知识图谱在智能电网的应用[J]. 电子元器件与信息技术, 2020, 4(1): 135-137, 147.
- [4] 刘峤, 李杨, 段宏, 刘瑶, 秦志光. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 13-16.
- [5] 王渊, 彭晨辉, 王志强, 等. 知识图谱在电网全业务统一数据中心的应用[J]. 计算机工程与应用, 2019, 55(15): 104-109.
- [6] 汤亚宸, 方定江, 韩海韵, 等. 基于图数据库和知识图谱的电力设备质量综合管理系统研究[J]. 供用电, 2019(11): 35-40.
- [7] 高泽璞, 赵云, 余伊兰, 等. 基于知识图谱的低压配电网拓扑结构辨识方法[J]. 电力系统保护与控制, 2020,

48(2): 34-43.

- [8] 严玉良, 董一鸿, 何贤芒, 等. FSMBUS: 一种基于 Spark 的大规模频繁子图挖掘算法[J]. 计算机研究与发展, 2015, 52(8): 1768-1783.
- [9] Yan, X. and Han, J. (2002) gSpan: Graph-Based Substructure Pattern Mining. *Proceedings of the IEEE International Conference on Data Mining*, Maebashi City, 9-12 December 2002, 721-724.
- [10] Elseidy, M., Abdelhamid, E., Skiadopoulou, S., et al. (2014) GRAMI: Frequent Subgraph and Pattern Mining in a Single Large Graph. *Proceedings of the Vldb Endowment*, 7, 517-528. <https://doi.org/10.14778/2732286.2732289>
- [11] Suchanek, F.M., Kasneci, G. and Weikum, G. (2007) Yago: A Core of Semantic Knowledge. *WWW'07: Proceedings of the 16th International Conference on World Wide Web*, May 2007, 697-706. <https://doi.org/10.1145/1242572.1242667>