

基于数据流的桥梁健康监测海量数据处理^①

王家伟, 汪仁红, 罗 宪, 黄 坤

(重庆交通大学 信息科学与工程学院, 重庆 400074)

摘 要: 采用数据流的滑动窗口技术, 建立了一种基于数据流滑动窗口的桥梁健康监测系统数据处理及存储模型, 进行了基于数据流的桥梁健康监测数据处理的研究。桥梁健康监测是对桥梁结构进行实时监控和结构状态的智能化评估。这对确保桥梁的运营安全及早发现桥梁病害, 对延长桥梁的使用寿命起着积极的作用。基于数据流的滑动窗口技术, 对传感器采集的数据进行实时处理, 并根据数据流设置的阈值对异常数据进行保存, 然后根据异常数据判断是否报警。该模型大大减少了数据处理的难度, 同时也提高了数据处理的效率。

关键词: 桥梁结构; 桥梁监测; 数据流; 滑动窗口; 数据处理

Processing of Massive Data Based on Data Stream and Bridge Health Monitoring

WANG Jia-Wei, WANG Ren-Hong, LUO Xian, HUANG Kun

(College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: This paper established a processing and storage model of a bridge health monitoring system based on data stream sliding window by using the data stream and the sliding window technical. It studied the bridge health monitoring of data processing based on data flow. Bridge health monitoring is a real-time monitoring for the bridge structure and the structural state of the intelligence assessment. It plays an active role for ensuring the safe operation of the bridge, detecting the diseases of the bridge and extending the life of the bridge. The data stream and sliding window technical take real-time processing to data which collected by the sensors, and set the threshold based on the data stream to save on the abnormal data, then according to the abnormal data determine whether alarm. This model greatly reduces the difficulty of data processing, while improving data processing efficiency.

Key word: bridge structure; bridge monitoring; data stream; sliding window; data processing

1 桥梁结构的健康监测的概况

桥梁健康监测是指运用现代化传感设备与光电通信技术以及计算机技术对桥梁相应关键部位, 实时的在线测量桥梁结构在运行过程中的各种结构响应参数, 并将这些数据传输给服务器中心。按照响应阈值以及评估方法及时的对桥梁结构的健康状况作出评估, 在出现异常时采取相应的处理措施或给出报警。一般桥梁健康监测对以下几方面进行监测: 桥梁在正常车辆荷载及风载作用下的结构响应和力学状态; 桥梁在突发事件(如强烈地震、意外大风或其它严重事故等)之后的剩余寿命评估; 大桥所处的环境条件, 如风

速、温度、地面运动; 大桥响应特性, 如应力、位移、动静态特性等。

桥梁结构的健康状况是由测试的数据来监测和评估的, 即从传感器采集的数据中提取各种特征, 对结构进行参数检测、状态监控和损伤诊断等。其中桥梁相应结构参数如动应变、振动、索力等属于变化周期较短的参数, 能实时反应桥梁结构的运营状况, 每天的纯数据就有上 G 的字节, 再加上数据之外的控制信息, 数据会达到几十 GB 级别。为了识别和定位损伤需要有效、可靠的数据分析与处理方法。而在目前已有的桥梁健康监测系统中, 在对监测数据的管理方面,

^① 基金项目:重庆市攻关项目(2010AC2077)

收稿时间:2011-04-09;收到修改稿时间:2011-05-06

还没有形成一个较为完善的数据存储与管理查询系统,大量的监测数据得不到妥善的处理和利用。本文基于数据流技术,结合数据流的滑动窗口,建立了一种基于数据流的桥梁健康监测系统的数据处理及存储模型。

2 数据流技术

2.1 数据流定义

数据流是一个由连续的、随时间变化的、无界的数据项构成的序列。流中的数据项的形式可能是关系元组,也可能是对象的实例。在大多数流式应用中,数据项是以关系元组的形式出现的,我们以此对基于关系的数据流定义如下:

一个由有先后顺序关系且个数随时间不断增加的元组构成的数据集。

形式化描述为: t_n , 其中, t 为元组 s 的时间戳, T 为数据流系统中的时间域。

从本定义可以看见此数据流具有无限性、广域性、高效性、动态性、瞬时性、时效性等特征。而数据流的这些特点决定了数据流的处理特点。如数据流的高速性,即在短时间内可能会有大量的输入数据需要处理,而这对处理器和输入输出设备来说都是一个较大的负担,因此对数据流的处理应尽可能简单;数据流的广域性,即数据的取值范围非常大,可能取的值非常多,这就导致数据流无法在内存或硬盘中存储,这就意味着系统无法完整保存这些数据,通常只能在数据到达的时候存取数据一次。如果范围小,即使到来的数据量很大,也可以在较小的存储器中保存这些数据;数据流的持续性,即数据的持续到达使数据量可能是无限的。而且对数据进行的处理结果不会是最终的结果,因为数据还会不断地到达。因此,对数据流的查询的结果往往不是一次性而是持续性的,即随着底层数据的到达而不断返回最新的结果。这些数据流的特点就决定了数据流处理的一次存取,持续处理,有限存储,近似结果,快速响应等特点。

近似结果是在前三个条件限制下产生的必然结果。由于只能存取数据一次,而且只有相对较小的有限空间存储数据,因此产生精确地计算结果通常是不可能的。而将对结果的要求从过去的“精确”改为“近似”后,实现数据流查询的快速响应也就成为了可能。

而我们在进行数据流处理时,有哪些数据是被包

含在处理范围内的,关于这个问题,本文主要有三种不同的模型:分别是数据流模型(Data Stream Model)、滑动窗口模型(Sliding Window Model)和 n -of- N 模型。

在数据流模型中,从某个特定时间开始至今的所有数据都要被纳入处理范围。此时 $s=0$,即在时刻 0, α 是 0 向量。这是数据流最初和最普遍的模式。

在 n -of- N 模型中,从计算时算起,计算最近的 n ($0 < n \leq N$) 个数据,即向前追溯的 n 个数据。此时, $s = t_n$,即在 t_n 时刻 α 是 0 向量。注意,其中 $n \leq N$,而且是可以随查询要求变化的。

滑动窗口模型是指,从计算时算起处理最近的 N 个数据,即向前追溯的 N 个数据要被纳入处理范围。此时, $s = t_n$,即在时刻 t_n , α 是 0 向量。由于数据流的数据是不断涌现的,所以直观的看,这种模式就像用一个不变的窗口,数据随时间的推移经过窗口,出现在窗口内的数据就是被处理的数据。

而本文建立了一种基于数据流的滑动窗口处理及存储模型。

2.2 滑动窗口模型

不同于传统的 DBMS,数据流管理系统(DSMS)处理数据时,必须具有—次存取、持续处理、有限存储、结果近似和快速响应的功能^[1]。DSMS 原理是建立一种存储机制,这种机制为人们提供一些重要的数据。这种机制可抽象成一种模型(如图 1 所示)。

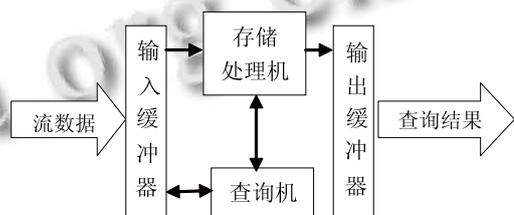


图 1 DSMS 模型

上图中,输入缓冲器的设置便于批量处理流数据,存储器处理机对窗口查询操作所需要的数据、流数据的统计信息以及根据特定的需求对某阶段的原始数据进行保存,查询处理机响应用户提出的查询请求,并对查询进行优化,最后由输出缓冲器输出查询结果。

在 DSMS 中可以采用滑动窗口技术来处理数据。存储处理的数据对象就是每一时刻流过这个滑动窗口的序列片段。滑动窗口中保存的是数据流中最近到达的那部分数据,随着新数据的不断到来,滑动窗口也

跟着不断更新。滑动窗口可分为基于元组的滑动窗口、基于时间的滑动窗口及分区窗口。基于元组的滑动窗口是大小固定的窗口,窗口模型以正整数 N 作为参数,只有当数据达到一定数目之后才能做相应的处理。基于时间的滑动窗口^[2]是大小可变的窗口,窗口模型以时间间隔 T 作为参数,为数据流运行时间的 T 计算周期。输出关系可定义为:

$$R(\tau) = \{S | \langle S, \tau' \rangle \in S \wedge (\tau' \leq \tau) \wedge (\tau' \geq \max\{\tau - T, 0\})\} \quad (1)$$

有两种特殊情况:

当 $T=0$ 时,由数据流 S 中带有时间戳的元素组成;

当 $T=\infty$ 时,由数据流 S 中所有时间戳的元素组成。

分区窗口是将数据流 S 按属性 A_1, \dots, A_k 的相等性(类似于分组)进行逻辑分区,分成各个不同的子数据流,然后对这些子数据流分别计算出大小为 N 的基于元组的滑动窗口。形式化来说,一个元组 S , 其属性值为 a_1, \dots, a_k 输出的瞬时关系 $R(x)$, 当且仅当存在一个元素 $\langle s, \tau' \rangle \in S, \tau' \leq T, \tau'$ 是那些属性值为 a_1, \dots, a_k 的 N 个元素中的最大时间戳^[3]。

3 基于数据流的桥梁健康监测数据处理模型

3.1 桥梁监测中的海量数据

桥梁健康监测基于大量的前端传感器实时监控采集桥梁相应结构参数,如形变、静应变、动应变、振动、索力等相关数据。形变和静应变属于变化周期较长的参数,因此数据测量周期可以是小时级别,数据量有限,采用传统的 DBMS 可以满足数据的存储和查询处理的要求。但动应变、振动、索力等属于变化周期较短的参数,要求 7×24 小时不间断工作,能实时反应桥梁的状况。以某桥为例,动态测点为 86 个,每个测点每秒钟采集一次数据,16 位的精确度,按照关系模式要求计算,每秒钟产生 86 条记录,每天就有近 800 万条记录。一天的纯数据就有 2G 多字节,再加上数据之外的控制信息,数据会达到几十 GB 级别。对于这样级别的海量数据若在同一台服务器上完成数据存储、检索服务及后期数据处理服务,传统意义上的数据库数据管理已经到了极限,传统的 DBMS 并不能承载这类级别的数据及处理结果。

3.2 基于滑动窗口的数据处理模型

通过分析发现,桥梁相应结构参数如动应变、振

动、索力等变化周期较短的监测数据元素的值是不确定的,数据的到达具有快速的、随时间变化的、无限的、不可预测等特点,不可能将数据流中所有的数据存储下来^[4],但存储单个数据结构体的大小是一致的,符合数据流的特征,也称得上是一种有序平稳的数据流^[5]。因此我们采用数据流技术,建立了基于滑动窗口的数据处理及存储模型,如图 2。由于数据流是稳定的,滑动窗口可以采用基于时间的,也可以采用基于数据元素序列的。

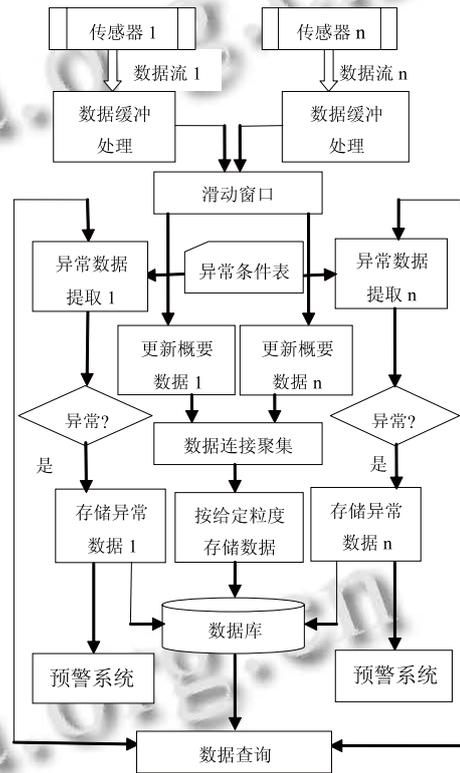


图 2 基于滑动窗口的数据处理及储存模型

每个传感器实时监测数据都形成一个数据流,数据缓冲处理用于接收数据流数据,同一时刻同一类型的流数据按传感器的位置编号进行分类,将缓冲区中的数据加入到滑动窗口,根据各数据流设置的阈值判断是否是异常数据,若是异常数据,则将该监测数据保存到数据库中,按照预警模型分析决定是否报警。

滑动窗口根据需要设置滑动周期,当到达滑动周期时间时,数据处理模块完成该周期内各数据流的概要数据的计算,更新当前概要数据结构,将该概要数据保存到数据库中,同时抛弃前一个周期的数据。

数据查询分为两种类型的查询:精确查询和近似

查询。精确数据查询是查询滑动窗口中的数据,即查询滑动窗口的一个周期内的数据;近似查询是根据聚集条件查询给定时间段内的概要数据。在本模型中,近似查询方式可以结合异常概要数据进行综合数据聚集查询。

3.3 数据存储量分析

数据存储量以关系模型的一个元组为单位,时间以秒为单位。设动态监测传感器的数量为 n , 动态监测传感器监测周期为 t_1 , 静态监测周期为 t_2 , 一个数据流在一个静态监测周期中的平均异常数据量为 m_1 , 每天存储的数据量为 M 。存储所有监测数据量为:

$$M = \frac{24 * 3600}{t_1} \times n \quad (2)$$

本模型监测数据量为:

$$M = \left(\frac{24 * 3600}{t_2} + m_1 \right) \times n \quad (3)$$

式(3)中 M 的值取决于 t_2 和 m_1 。一般来说,在实际的监测应用中, t_2 远大于 t_1 。在各数据流的阈值设置合理的情况下, m_1 的量较小。

以某新建桥梁为例,动态测点有 86 个,每个测点的监测周期为 1s,静态监测周期为 1h。若存储所有监测数据,则每天的元组数为 7400000。由于该桥是新建桥梁,使用本模型完成数据处理和存储,观测多天的数据发现, m_1 的数量趋近于 0,实际每天存储的数据为 2000 多个元组。

4 结语

本文围绕桥梁健康监测中的海量数据处理问题,提出并建立了一种基于滑动窗口的数据处理及存储模型。该模型应用于桥梁健康监测工程中是可行的,它对传感器采集的数据进行实时处理,根据数据流设置的阈值对异常数据进行保存,然后根据异常数据判断是否报警。数据流经过滑动窗口根据计算得到一个周

期内的概要数据,并将得到的概要数据存储到数据库。从而大大减少了数据处理的量度以及难度,也提高了数据处理的效率。

参考文献

- 1 李子杰,郑成.流数据和传统数据存储及管理方法比较研究.计算机技术与发展,2009,19(4):101-104.
- 2 杜威,邹先霞.基于数据流的滑动窗口机制的研究.计算机工程与设计,2005,26(11):2922-2924.
- 3 李琳.环境监测中的数据流处理技术研究及实现.国防科技大学,2007.
- 4 王伟平,李建中,张冬冬,郭龙江.数据流上周期更新滑动窗口的连接算法.哈尔滨工业大学学报,2005,37(6):756-759.
- 5 刘欣,李永刚.海量数据流在桥梁健康监测中的处理方法.计算机工程与设计,2008,29(1):223-238.
- 6 张驹,黄汉永,肖杰.一种基于 Hash 函数抽样的数据流聚类算法.计算机系统应用,2009,18(3):73-75.
- 7 张龙波,李战怀,余敏,蒋芸.带权值数据流滑动窗口随机抽样算法的改进.计算机工程与应用,2007,43(25):18-20.
- 8 刘佳,刘国华,宋驰.基于流数据技术的连续查询处理.计算机工程,2005,31(8):71-73.
- 9 王翔,李建中,王伟平.基于滑动窗口的数据流压缩技术及连续查询处理方法.计算机研究与发现,2004,41(10):1639-1644.
- 10 Datar M, Gionis A, Indyk P, et al. Maintaining stream statistics over sliding windows. SIAM Journal on Computing, 2001,31:1794-1813.
- 11 王永利,董逸生,徐宏炳,钱江波,刘学军.数据流关键技术研究.计算机时代,2004,(12):6-10.
- 12 钱江波,徐宏炳,王永利,刘学军,董逸生.多数据流滑动窗口并发连接方法.计算机研究与发展,2005,42(10):1771-1778.
- 13 宫学庆,闫莺,常建龙,张晨,周傲英.数据流处理技术在电网管系统中的应用.计算机科学与探索,2008,(2):1673-9418.