



(12)发明专利申请

(10)申请公布号 CN 109542803 A

(43)申请公布日 2019. 03. 29

(21)申请号 201811385060.4

(22)申请日 2018.11.20

(71)申请人 中国石油大学(华东)

地址 266580 山东省青岛市黄地区长江西路66号

(72)发明人 张卫山 房凯 任鹏程

(51)Int.Cl.

G06F 12/0895(2016.01)

权利要求书1页 说明书3页 附图2页

(54)发明名称

一种基于深度学习的混合多模式热数据缓存策略

(57)摘要

本发明提出一种基于深度学习的混合多模式热数据缓存策略。在索引缓存层快速响应大部分的查询请求,减少持久存储层的访问次数,提高索引系统的整体性能是分层式索引模型的设计目标。但缓存的容量毕竟是有限的,如何选择一种合适的缓存替换策略,在使用尽量少的缓存情况下,提高其命中率成为提升索引系统整体性能的一个重要研究问题,因此我们提出基于深度学习的混合多模式热数据缓存替换策略。在原有的多模式热数据敏感的缓存策略基础上,通过支撑平台提供的查询趋势预测服务,对查询任务进行预测,采用深度学习中的DBN预测算法来预测在完成一次查询任务后,下一次可能的查询任务,将与预测到的查询任务相关的索引记录加载至缓存层中,从而增大缓存中索引记录的命中率,提高查询效率。



1. 一种基于深度学习的混合多模式热数据缓存策略,其特征在於,在原有的多模式热数据敏感的缓存策略基础上,采用深度学习中的DBN预测算法作为多模式热数据敏感的缓存替换策略的补充和加强,包括以下步骤:

步骤(1)、聚合数据;

步骤(2)、对(1)的数据进行差分变换以降低数据的线性度,归一化数据,分析数据内部的自相关和自回归特性;

步骤(3)、按照从下到上的顺序逐层训练RBM,并通过BP算法优化整体网络结构;

步骤(4)、使用前一步训练出的网络对云属性进行长短期预测;

步骤(5)、将(4)中的预测结果进行与(2)相反的逆变换,得到原数据的预测值。

一种基于深度学习的混合多模式热数据缓存策略

技术领域

[0001] 本发明涉及缓存替换策略,智能预测和深度学习,具体涉及到一种基于深度学习的智能预测方法。

背景技术

[0002] 通过在索引缓存层快速响应大部分的查询请求,以减少持久存储层的访问次数,从而提高索引系统的整体性能。由于缓存的容量是有限的,需要选择一种合适的缓存替换策略,在使用尽量少的缓存情况下,提高其命中率,提升索引系统的整体性能。在内存索引服务进程中,使频繁访问的数据能够贮存在索引缓存层,从而更有效地利用内存空间并提升索引查询性能。最接近本发明的技术有:

[0003] (1)、最不经常使用(LFU)算法:LFU(Least Frequently Used,最不经常使用)算法将一段时间内被访问次数最少的那个块替换出去。每块设置一个计数器,从0开始计数,每访问一次,被访块的计数器就增1。当需要替换时,将计数值最小的块换出,同时将所有块的计数器都清零。这种算法将计数周期限定在对这些特定块两次替换之间的间隔时间内,不能严格反映近期访问情况,新调入的块很容易被替换出去。

[0004] (2)、近期最少使用(LRU)算法:LRU(Least Recently Used,近期最少使用)算法是把CPU近期最少使用的块替换出去。这种替换方法需要随时记录Cache中各块的使用情况,以便确定哪个块是近期最少使用的块。每块也设置一个计数器,Cache每命中一次,命中块计数器清零,其他各块计数器增1。当需要替换时,将计数值最大的块换出。LRU算法相对合理,但实现起来比较复杂,系统开销较大。这种算法保护了刚调入Cache的新数据块,具有较高的命中率。LRU算法不能肯定调出去的块近期不会再被使用,所以这种替换算法不能算作最合理、最优秀的算法。但是研究表明,采用这种算法可使Cache的命中率达到90%左右。

[0005] (3)、随机替换:随机替换算法完全不管Cache的情况,简单地根据一个随机数选择一块替换出去。随机替换算法在硬件上容易实现,且速度也比前两种算法快。缺点则是降低了命中率和Cache工作效率。

[0006] (4)、基于近期最少使用(LRU)算法的缓存替换策略:使用指数平滑方法来预测索引记录的热度,在替换时根据预测热度值来决定当前替换的内容。首先,按照热度预测公式对所有的索引记录进行热度预测。然后根据预测热度值对记录进行排序。最后按预测热度值从高到低将索引记录缓存到内存中,直到内存缓存用满为止。

[0007] 为了能够更加精准的选择加载至缓存中的索引,在原有的多模式热数据敏感的缓存策略基础上,通过支撑平台提供的查询趋势预测服务,对查询任务进行预测。在查询趋势预测服务中,采用深度学习中的DBN预测算法来预测在完成一次查询任务后,下一次可能的查询任务,将与预测到的查询任务相关的索引记录加载至缓存层中,从而增大缓存中索引记录的命中率,提高查询效率,作为多模式热数据敏感的缓存替换策略的补充和加强。

发明内容

[0008] 为解决现有技术中的缺点和不足,本发明提出了基于深度学习的混合多模式热数据缓存策略,采用深度学习中的DBN预测算法作为多模式热数据敏感的缓存替换策略的补充。

[0009] 本发明的技术方案为:

[0010] 一种基于深度学习的混合多模式热数据缓存策略,其特征在于,在原有的多模式热数据敏感的缓存策略基础上,采用深度学习中的DBN预测算法作为多模式热数据敏感的缓存替换策略的补充和加强,包括以下步骤:

[0011] 步骤(1)、聚合数据;

[0012] 步骤(2)、对(1)的数据进行差分变换以降低数据的线性度,归一化数据,分析数据内部的自相关和自回归特性;

[0013] 步骤(3)、按照从下到上的顺序逐层训练RBM,并通过BP算法优化整体网络结构;

[0014] 步骤(4)、使用前一步训练出的网络对云属性进行长短期预测;

[0015] 步骤(5)、将(4)中的预测结果进行与(2)相反的逆变换,得到原数据的预测值。

[0016] 本发明的有益效果:

[0017] (1)在索引缓存层快速响应大部分的查询请求,减少持久存储层的访问次数,提高索引系统的整体性能;

[0018] (2)在使用尽量少的缓存情况下,提高其命中率;

[0019] (3)在内存索引服务进程中,使频繁访问的数据能够贮存在索引缓存层,从而更有效地利用内存空间并提升索引查询性能;

[0020] (4)采用深度学习中的DBN预测算法来预测在完成一次查询任务后,下一次可能的查询任务,将与预测到的查询任务相关的索引记录加载至缓存层中,从而增大缓存中索引记录的命中率,提高查询效率。

附图说明

[0021] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0022] 图1为本发明基于深度学习中DBN算法的模型图。

[0023] 图2为本发明基于DBN算法的智能预测过程。

具体实施方式

[0024] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0025] 如图1所示,基于深度学习DBN算法的智能预测方法的过程主要包括3个部分:对数据进行预处理、DBN预训练和BP算法的优化和预测结果的后处理。

[0026] 下面结合图1与图2,对基于深度学习的智能预测方法的具体流程进行详细说明:

[0027] 步骤(1)、聚合数据;

[0028] 步骤(2)、对(1)的数据进行差分变换以降低数据的线性度,归一化数据,分析数据内部的自相关和自回归特性;

[0029] 步骤(3)、按照从下到上的顺序逐层训练RBM,并通过BP算法优化整体网络结构;

[0030] 步骤(4)、使用前一步训练出的网络对云属性进行长短期预测;

[0031] 步骤(5)、将(4)中的预测结果进行与(2)相反的逆变换,得到原数据的预测值。

[0032] 本发明的基于深度学习的热数据敏感的缓存替换策略,可以预测数据的使用情况,并按热度由高到低确定保留在内存中索引,能较好地适应不同的应用场景并有效地将热数据缓存在内存中,从而进一步提高了分层式索引系统的查询效率。

[0033] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。



图1

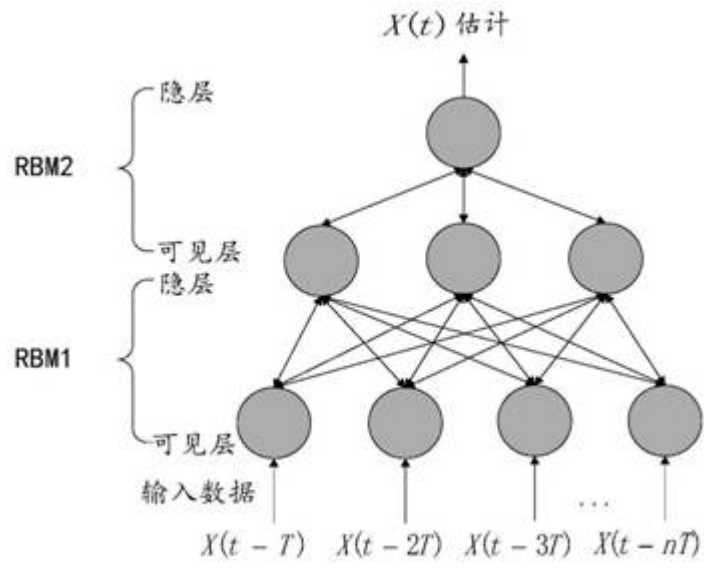


图2