

多干扰网络的海量数据优化调度模型仿真分析

武怀生

(青海民族大学 物理与电子信息工程学院, 青海 西宁 810007)

摘 要: 网络中存在很多干扰信息,且数据节点的利用有很强的随机性,当前数据调度模型容易造成数据节点冲突,导致调度效率降低,为此,提出一种新的多干扰网络的海量数据优化调度模型,构建网络待处理信息特征相似性模型,获取网络待处理项的特征邻居,采用协同过滤算法实现网络多干扰过滤.在此基础上对海量数据调度任务量进行分区计算,获取海量数据调度任务中的子任务,从而构建海量数据调度任务和数据节点间的映射关系.依据目标数据在服务器的传输状态求出海量数据通信的任务量,构建海量数据优化调度模型.实验结果表明,采用所提模型对海量数据进行调度,不仅吞吐量高,而且传输速率快,具有很高的调度性能.

关键词: 多干扰网络;海量数据;调度模型

中图分类号: TP391.9

文献标识码: A

文章编号: 1000—7180(2016)09—0155—04

Multi-Interference Network Optimization Scheduling Model Simulation Analysis of Huge Amounts of Data

WU Huai-sheng

(School of Physics and Electronic Information Engineering, Qinghai University of Nationalities, Xining 810007, China)

Abstract: There's so much interference in the network information and the use of data nodes have strong randomness, the current data scheduling model is easy to cause conflict data node, results in the decrease of scheduling efficiency, therefore, put forward a new network more interference of huge amounts of data optimization scheduling model, build the network information to be processed characteristic similarity model, the characteristics of the item to be processed to obtain network neighbors, the collaborative filtering algorithm was adopted to realize the network interference filter. Based on the partition for the amount of mass data scheduling tasks is calculated, to obtain huge amounts of data scheduling tasks of subtasks, so as to build huge amounts of data scheduling tasks and data mapping relationships between the nodes. According to the target data transmission in the server's status, the quota of huge amounts of data communication, build huge amounts of data optimization scheduling model. The experimental results show that the proposed model to the scheduling of huge amounts of data, not only high throughput, and transmission rate is fast, has a high operation performance.

Key words: network more interference; huge amounts of data; scheduling model

1 引言

随着计算机和网络的逐渐发展,互联网已经成为应用最广泛的信息网络.网络在给人们的生活带来便利的同时也隐藏很多干扰信息^[1-3].在这种情况下滤除干扰和数据管理已成为多个领域的核心^[4,5].因此,如何针对海量数据进行调度,已经成

为数据管理领域学者研究的重点课题^[6].

文献[7]将丢包率作为目标对海量数据进行调度,从接收节点利益最大化的角度实现数据调度.该方法能够使自身利益达到最大,但会导致部分传输效率高的节点持续传输,负载较大,降低其参与传输的积极性,影响传输效率;文献[8]提出一种基于蚁群算法的海量数据调度模型,在对海量数据进行调

度时,可在庞大的解空间中找到全局最优解,实现海量数据的优化调度.该方法在解决组合优化问题方面性能较高,然而其具有初期收敛速度慢,易陷于局部最优解的弊端;文献[9]提出一种集中式海量数据聚集调度模型,通过已知节点获取网络的全部信息,通过获取的信息制定调度方案,发送至各节点中.该方法调度性能较高,但仅在全部节点接收到调度方案后才可实施调度,若网络拓补结构出现改变,节点需重新制定调度方案并传输至各节点,导致能耗高,效率低;文献[10]提出一种可感知链路质量的数据调度模型,发送端对不同路径的网络状态进行评估,依据评估结果对各待传输数据包的传输路径与时间进行设置,从而实现海量数据的优化调度.该方法调度效率高,但忽略了不同路径带宽、时延上的差异,无法有效利用网络状况较优的路径.

针对上述方法的弊端,提出一种新的多干扰网络的海量数据优化调度模型,采用协同过滤算法实现网络多干扰过滤.在此基础上对海量数据调度任务量进行分区计算,从而构建海量数据优化调度模型.实验结果表明,采用所提模型对海量数据进行调度,不仅吞吐量高,而且传输速率快,具有很高的调度性能.

2 多干扰网络的海量数据优化调度模型

2.1 网络多干扰过滤

网络中的干扰信息对网络性能产生了不好的影响,使网络运行效率大大降低.因为网络多干扰信息量较大,所处环境噪声较多,采用传统方法实现网络多干扰过滤时,通常过滤不充分,因此,本节采用协同过滤算法实现网络多干扰过滤.

2.1.1 构建网络待处理信息特征相似性模型

本节通过网络特征数据量化、相似性计算与特征邻居搜索对网络待处理信息特征相似性模型进行建立,详细建立过程如下所述.

假设由 n 个网络待处理信息属性构成的 n 维向量用 $(I_{i1}, I_{i2}, \dots, I_{in})$ 进行描述,其中, I_{in} 用于描述包含在第 i 个项目中的第 n 个属性的特征值.

为了不忽略维度较小的待处理信息的特征向量的计算,通过下式求出特征相似度:

$$s(i, j) = \frac{|N_{ij}|_{t_{ic}^{=1} \cap t_{jc}^{=1}}}{|N_{all}|_C - |N_{ij}|_{t_{ic}^{=0} \cap t_{jc}^{=0}}} \quad (1)$$

式中, $|N_{ij}|_{t_{ic}^{=1} \cap t_{jc}^{=1}}$ 表示两个项目具有的相同的特征数; $|N_{all}|_C$ 表示特征数量; $|N_{ij}|_{t_{ic}^{=0} \cap t_{jc}^{=0}}$ 表示不含任何一个项目的特征数.

通过计算可获取下述相似性矩阵:

$$\begin{matrix} I_{i1} & I_{i2} & I_{i3} & \cdots & I_{in} \\ \begin{bmatrix} 1 & & & & \\ S_{21} & 1 & & & \\ S_{31} & S_{32} & 1 & & \\ \cdots & \cdots & \cdots & 1 & \\ S_{n1} & S_{n2} & \cdots & S_{n(n-1)} & 1 \end{bmatrix} \end{matrix} \quad (2)$$

为了增强项目的查询速率可构建图 1 所示的特征邻居排序链.

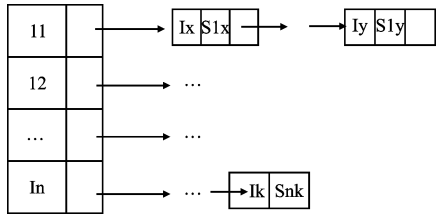


图 1 特征邻居排序链

图 1 中的相似度是按照降序排列的, I_i 为项目, S_{ij} 描述了项目 I_i 与 I_j 的相似度.

通过上述过程可建立网络待处理项的特征相似性模型.

2.1.2 获取网络待处理项的特征邻居

获取网络待处理项的特征邻居对网络多干扰过滤方法的准确性起到了至关重要的作用,下面进行详细的分析.

假设需求出干扰项目 i 和 j 之间的相似性,首先对评价过 i, j 的全部用户进行统计:

$$U_{ij} = \{u | u \in U \cap (r_{u,i} \neq 0 \text{ OR } P_{u,i} \neq 0)\} \quad (3)$$

式中, u 代表用户; U 代表用户集合; $r_{u,i}$ 与 $P_{u,i}$ 代表评分结果.通过特征相似群的评分对没有被用户评价的项目进行预评分,通过以上预评分结果,依据 Pearson 相关系数法求出 i 和 j 的相似性,从而获取目标干扰项的最近邻居.其中预评分可利用下述过程求出:

已知集合 U_{ij} 中评价过 i 的用户 u 对其的评分如下:

$$R_{u,i} = \begin{cases} r_{u,i} & r_{u,i} \neq 0 \\ P_{u,i} & r_{u,i} = 0 \end{cases} \quad (4)$$

如果用户 u 已经对该项目进行了评价,则其评分结果是 $r_{u,i}$;如果用户 u 没有对该项目进行评价,则可通过特征邻居对评分进行预测:

$$P_{u,i} = \frac{\sum_{j \in C_j} s(i, j) \times r_{u,j}}{\sum_{j \in C_j} |s(i, j)|} \quad (5)$$

式中, C_j 用于描述项目 j 的最近特征邻居.

2.1.3 获取过滤结果

依据上述分析可知,如果假设网络用户 A 的待处理目标项 TI 的特征邻居集合是 $NN_{TI} = \{NN_1, NN_2, \dots, NN_k\}$, 则用户 A 对 TI 的预评分结果可描述成:

$$P_{A, TI} = \overline{R}_{TI} \frac{\sum_{j \in NN_{TI}} \text{sim}(TI, j) \times R_{u, j} - \overline{R}_j}{\sum_{j \in NN_{TI}} |\text{sim}(TI, j)|} \quad (6)$$

式中, $\text{sim}(TI, j)$ 用于描述待测目标项 TI 预期最近邻居 j 的相似性; $P_{A, j}$ 用于描述用户 A 对 j 的评分结果.

通过上述分析对网络待处理项的预测评分进行排列, 选择排列顺序靠前的 N 个项目为用户提供服务, 将剩下的看作干扰项进行滤除.

2.2 海量数据优化调度模型

2.2.1 分区计算调度任务量

完成网络多干扰滤除后, 对海量数据调度任务量进行分区计算.

在对海量数据进行优化调度时, 需将所有数据调度任务分割为多个子任务, 用 $c = [c(0), c(1), \dots, c(e-1)]^k$ 进行描述, 其分布符合 $a \sim g(c, s)$ 条件.

通过下式求出总调度任务量:

$$\hat{s} = \arcsin \min_{q \in \epsilon} \sum_{s=1}^{+\infty} H(\hat{s}, s) g(s|c) \quad (7)$$

式中, $H(\hat{s}, s)$ 用于描述调度任务中代价函数; $g(s|c)$ 用于描述调度任务中子任务的分布概率; ϵ 用于描述任务调度中全部子任务的集合, 其可通过下式求出:

$$\epsilon = \{\hat{s} | c_1 + 2c_r \leq \hat{s} \leq S\} \quad (8)$$

式中, S 用于描述调度任务过程中的任务量极大值. 通过下式求出其代价函数:

$$H(\hat{s}, s) = (\hat{s} - s)^2 \quad (9)$$

将式(9)代入式(7)中, 可获取以下结果:

$$\hat{s} = \sum_{s=1}^S s \hat{s}(s|c) \quad (10)$$

式中, $\hat{s}$ 用于描述调度过程所需的时间均值. 假设调度时间符合 $\sum_{s=1}^S \hat{s}(s|c) = 1$ 的条件, 则可获取估计函数:

$$H(\hat{s}, s) = |\hat{s} - s| \quad (11)$$

将上式代入调度任务量计算公式中, 从而获取以下结果:

$$\hat{s} = \arcsin \min_s \left(\sum_{s=1}^{\hat{s}} s(s|s) - \sum_{s=\hat{s}}^s s(s|c) \right) \quad (12)$$

通过下式可求出调度过程中的子任务量:

$$s = \arcsin \max_q (s|c) \quad (13)$$

通过上述分析可获取海量数据调度任务中的子任务, 为海量数据优化调度模型的构建提供准确数据基础.

2.2.2 构建调度任务优化模型

网络中存在很多干扰信息, 且数据节点的利用有很强的随机性, 当前数据调度模型容易造成数据节点冲突, 导致调度效率降低, 为此, 提出一种新的多干扰网络的海量数据优化调度模型.

假设数据节点单位时间发送的数据量为 z , 数据节点发送效率的模糊值为 θ , $\theta \in [0, 1]$, 通过下式可求出数据节点的发送效率递增系数:

$$\theta = v(z) \quad (14)$$

通过上述分析可构建海量数据调度任务和数据节点间的映射关系. 在上述映射关系中, 假设由所有调度任务量组成的数据集为 $\theta = \{\theta_1, \theta_2, \dots, \theta_q\}$, 其中 θ_i 用于描述第 i 个数据节点所需发送的数据. 假设 $j = \{\sigma_1 j_1, \sigma_2 j_2, \dots, \sigma_q j_q\}$, 其中 j_i 用于描述目标数据针对数据节点 T 的发送要求, 同时需符合 $U_i, j_i \in [0, 1]$ 的条件; σ_i 用于描述海量数据在调度时的优先级别, $\sigma_i \in [0, 1]$. 数据的优先级别越高, 越需对其进行优先调度. 在对任务进行时, 数据节点可发送目标数据的量是有限的, 需满足下式:

$$h(\theta_1, \theta_2, \dots, \theta_q) = 1 \quad (15)$$

通过下式可求出调度任务效率:

$$\varphi = |\theta - j| = \sqrt{(\theta_1 - \sigma_1 j_1)^2 + (\theta_2 - \sigma_2 j_2)^2 + \dots + (\theta_q - \sigma_q j_q)^2} \quad (16)$$

式中, φ 用于描述调度任务的效率; θ 和 j 的关联性越小, 则调度任务效率越高.

通过下式可构建海量数据调度优化模型:

$$\min \varphi = \sqrt{(\theta_1 - \sigma_1 j_1)^2 + (\theta_2 - \sigma_2 j_2)^2 + \dots + (\theta_q - \sigma_q j_q)^2} \quad (17)$$

式中, $\theta_i, j_i, \sigma_i \in [0, 1], i = 1, 2, \dots, q$.

通过上述过程, 可在发送目标数据时, 依据目标数据在服务器的传输状态求出海量数据通信的任务量. 依据得到的结果, 构建海量数据优化调度模型, 从而防止目标数据在同一数据节点出现冲突.

3 仿真实验分析

为了验证本文提出方法的有效性,需要进行相关的实验分析.

将 NS-2.34 网络仿真工具作为本文仿真工具,网络拓扑用图 2 进行描述,发送端与接收端分别含有三个网络接口,假设路径 1 的带宽是 0.3 Mb/s,路径 2 的带宽是 0.6 Mb/s,路径 3 的带宽是 0.9 Mb/s,发送端应用层的数据发送速率是 1.4 Mb/s.

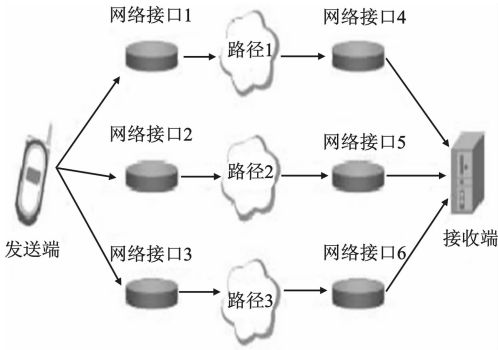


图 2 网络拓扑图

3.1 数据吞吐量仿真分析

吞吐量是指对网络、设备、端口、虚电路或其他设施,单位时间内成功地传送数据的数量.吞吐量越高,说明调度性能越高.图 3 描述的是采用本文方法前网络不同路径的平均吞吐量.

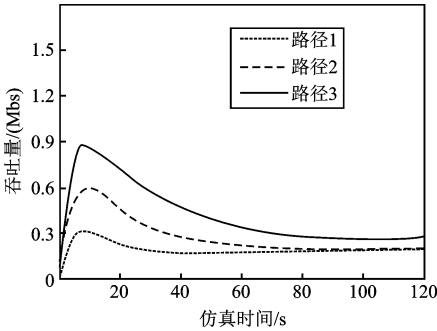


图 3 采用本文方法前网络不同路径的平均吞吐量

分析图 3 可知,在初始阶段,三条路径的数据吞吐量均逐渐增加;一段时间后,路径 2 与路径 3 的数据吞吐量开始迅速降低.这是因为接收缓存中阻塞了大量乱序数据包,导致接收缓存空闲减少,传输数据量变小.

图 4 描述了采用本文方法后网络不同路径的平均吞吐量.

分析图 4 可知,在曲线达到稳定状态后,三条路径的数据吞吐量均与其可用带宽相等,说明本文方

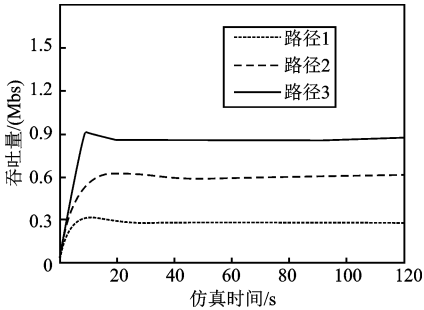


图 4 采用本文方法后网络不同路径的平均吞吐量

法能够使数据吞吐量达到最大,本文方法的带宽利用率很高.

3.2 数据传输效率仿真分析

在干扰网络环境下,不同路径的可用带宽与时延变化频率较高,路径中可用带宽容易受到很多因素的干扰.所以,一个性能优异的调度模型需适应不同网络环境,在网络环境不稳定的情况下仍可获取较高的传输效率.下面以带宽变化为例,对采样本文方法前后数据传输效率进行比较,得到的结果用表 1 进行描述.

表 1 数据传输效率比较结果

带宽/(Mb/s)				传输时间/s	
路径 1	路径 2	带宽之差	带宽之比	采用本文方法前	采用本文方法后
0	1.4	1.4	0.00	10.78	10.78
0.1	1.3	1.2	0.08	34.73	10.86
0.2	1.2	1	0.17	22.64	10.85
0.3	1.1	0.8	0.27	17.33	10.82
0.4	1	0.6	0.40	13.77	10.74
0.5	0.9	0.4	0.56	12.26	10.75
0.6	0.8	0.2	0.75	10.84	10.89
0.7	0.7	0	1.00	10.11	11.03

分析表 1 可以看出,采用本文方法后的数据传输时间明显高于采样本文方法前,且随着路径带宽之差的改变,采样本文方法后的数据传输效率基本能保持稳定,说明本文方法能够有效适应带宽的变化,且数据传输效率很高.

4 结束语

本文提出一种新的多干扰网络的海量数据优化调度模型,构建网络待处理信息特征相似性模型,获取网络待处理项的特征邻居,采用协同过滤算法实现网络多干扰过滤.在此基础上对海量数据调度任

4 结束语

本文将深度信念网应用于模拟电路的故障诊断中,设计了两个实验,实验1在Leapfrog滤波器仿真电路上对比了深度信念网(DBN),单隐层BP网络及三隐层BP网络的故障诊断正确率,结果证明DBN在模拟电路故障诊断中的应用是可行的,实验2在带通电路上对比了DBN,单隐层BP和小波分析方法的特征提取能力,结果表明DBN特征提取能力优于其他两种方法,而且提取的特征更能反映数据本质.深度学习给模拟电路故障诊断提供了一个优质选择.

参考文献:

- [1] 孟凡斌,唐圣学,谭子尤. 基于小波和主元分析的容差电路的故障诊断方法[J]. 微电子学与计算机,2010,21(11):135-138.
- [2] 刘美容,张立玮. 基于小波分解和模糊聚类的模拟电路软故障诊断[J]. 微电子学与计算机,2014,31(12):140-143.
- [3] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786):504-507.
- [4] Hinton G E. A practical guide to training restricted boltzmann machines[M]. Neural Networks: Tricks of

the Trade of the series Lecture Notes in Computer Science,2012,9(1):599-619.

- [5] Hinton G E. Training products of experts by minimizing contrastive divergence[J]. Neural Computation, 2002, 14(8):1771-1800.
- [6] Geoffrey H. Where do features come from? [J]. Cognitive Science A Multidisciplinary Journal, 2013, 38(6):1078-1101.
- [7] Bengio Y. Learning deep architectures for AI[J]. Foundations & Trends in Machine Learning, 2009, 2(1):1-127.
- [8] 王承,陈光福,谢永乐. 小波-神经网络在模拟电路故障诊断中的应用[J]. 系统仿真学报,2005,17(8):1936-1938.
- [9] Zhang L, Sun L, Cai X, et al. Analog circuit fault diagnosis based on the wavelet analysis and EM algorithm of limited gaussian mixture model[J]. International Journal of Advancements in Computing Technology, 2013,10(5):32-42.

作者简介:

颜学龙 男,(1962-),教授,硕士研究生导师.研究方向为可测性设计、故障诊断和边界扫描技术.

邵伟(通讯作者) 男,(1987-),硕士研究生.研究方向为机器学习和故障诊断. E-mail:592226162@qq.com

(上接第158页)

务量进行分区计算,获取海量数据调度任务中的子任务,从而构建海量数据调度任务和数据节点间的映射关系.依据目标数据在服务器的传输状态求出海量数据通信的任务量,构建海量数据优化调度模型.实验结果表明,采用所提模型对海量数据进行调度,不仅吞吐量高,而且传输速率快,具有很高的调度性能.

参考文献:

- [1] 王菲.无线传感网络低空干扰下的通信优化模型仿真[J]. 计算机仿真,2015,32(1):331-334.
- [2] 孙卫真,王秀锦,徐远超. 交通信息分布式处理中的Hadoop调度算法优化[J]. 计算机工程与设计,2014,35(4):1269-1273.
- [3] 贾江涛. 基于动态规划框架下的水电系统优化调度[J]. 西安工程大学学报,2015,29(4):420-425.
- [4] 刘彤. 海量非线性网络流量数据分类模型的仿真分析[J]. 计算机仿真,2015,32(12):255-258.

- [5] 陈丽平,郭鑫. 最小能耗优化云模型中的动态图挖掘方法[J]. 计算机工程,2015,41(8):23-31.

- [6] 黄少卿,胡立强. 基于改进SALS算法的大数据挖掘效率优化探究[J]. 微型机与应用,2015,34(12):11-13.

- [7] 王艳. 云计算环境下差异性资源调度特征匹配算法[J]. 中国西部科技,2015,14(12):82-84.

- [8] 申倩,许美玉,姜春茂. 云计算环境下任务调度研究综述[J]. 智能计算机与应用,2014,4(6):75-77.

- [9] 熊最,王可人,金虎,等. 一种基于干扰对齐的多跳网络吞吐量优化模型[J]. 系统仿真学报,2015,27(6):1240-1246.

- [10] 何鹏举,吴来斌,宋凯华,等. 基才粒子群算法的嵌入式云计算资源调度[J]. 电子设计工程,2014,22(10):88-90.

作者简介:

武怀生 男,(1977-),硕士研究生,讲师.研究方向为信息技术与教育. E-mail:wuhuaisheng765@163.com