

文章编号: 2096-1472(2017)-09-21-03

弱关联冗余环境下的挖掘算法研究

蔡柳萍

(广东技术师范学院天河学院信息与传媒学院, 广东 广州 510540)

摘 要: 在弱关联冗余环境下, 开展的挖掘算法应用需要考虑关联属性, 本文主要从模糊神经网络学习算法与弱关联规则模型, 建立两方面内容展开探讨, 整理出算法应用所产生的数据信息, 对大数据环境挖掘算法进行深入研究, 为网络环境性信息处理任务的高效开展, 建立一个稳定基础环境。

关键词: 弱关联; 冗余环境; 挖掘算法

中图分类号: TP3-0 **文献标识码:** A

Research on the Mining Algorithm in the Weak Correlation Redundant Environment

CAI Liuping

(Tianhe College of Guangdong Polytechnic Normal University, School of Communication and Information, Guangzhou 510540, China)

Abstract: In the weak correlation redundant environment, the correlation attributes need to be taken into consideration in the application of mining algorithm. On the basis of the fuzzy neural network learning algorithm and the weak correlation rule model, the paper discusses the two aspects, sorts out the data generated through the algorithm, further studies the mining algorithm of big data environmental information, and establishes a stable fundamental environment in order to efficient process the network environment information.

Keywords: weak correlation; redundant environment; mining algorithm

1 引言(Introduction)

本文主要从模糊神经网络学习算法与弱关联规则模型建立两方面内容展开探讨, 整理出算法应用所产生的数据信息, 对大数据环境挖掘算法进行深入研究, 为网络环境性信息处理任务高效开展建立一个稳定基础环境。本文提出了一种基于弱聚类算法的云计算环境下海量数据中弱关联挖掘方法, 针对弱关联冗余环境进行, 通过数据的描述特征对数据特征进行分解, 依据数据特征, 对全部数据进行融合。基于关联决策概率将云计算环境下海量数据进行有效划分, 完成所有数据特征关联概率的计算, 通过弱聚类方法对属性元素进行分类, 将数量型元素转换成类别型, 通过弱化关联规则方法, 对经聚类处理后的数据进行挖掘。

2 弱关联冗余环境下的挖掘算法综述(Summarize on mining algorithm in weak correlation redundant environment)

2.1 模糊学习算法

模糊神经学习算法是当前数据挖掘的主流技术, 采用弱关联思想设计原则中, 系统之间关联会尽可能减少, 模拟学习算法应用后系统运算速度提升有明显帮助。弱关联环境下对信息数据处理使用效率都有极高的要求, 如果不能解决所遇到的问题, 在建立系统管理控制环境期间, 应用模糊数据运算分析学习方法, 能够帮助节省大量数据运算所用时间, 并在最终的控制计划由模糊学习算法中模拟神经元原理, 能够根据弱关联环境下的系统联系选择运输, 基于云计算环境

下不断学习数据更新, 达到记忆效果。模糊学习运算能够实现同步数据挖掘, 减少数据分析中不必要的时间。建立在弱关联环境中的各项学习控制计划中, 通过分析运算控制方法, 最终问题解决能力也不会因此受到影响。模糊学习算法是针对系统运行中控制能力提升进行的, 弱关联冗余系统部分在控制运算中通过学习算法可以将不必要的部分选择删除, 节省挖掘算法运行所用时间。模糊学习算法中会涉及隐藏数据, 模拟神经元来快速判断这部分信息^[1]。

2.2 挖掘算法中的决策树算法

应用该种算法, 需要选择属性用信息增益变化数据集, 建立符合运行模式的信息获取环境, 从而实现对信息传输数据决策范围判断, 也就是熵的变化值, 而C4.5用的是信息增益率, 也就是多了个率嘛。一般来说, 率就是用来取平衡用的, 就像方差起的作用差不多, 比如有两个跑步的人, 一个起点是100m/s的人、其1s后为110m/s; 另一个人起速是1m/s、其1s后为11m/s。在这里, 其克服了用信息增益选择属性时偏向选择取值多的属性的不足。在树构造过程中进行剪枝, 我在构造决策树的时候好讨厌那些挂着几个元素的节点。对于这种节点, 干脆不考虑最好, 不然很容易导致overfitting。对非离散数据都能处理, 这个其实就是一个个式, 看对于连续型的值在哪里分裂好。也就是把连续性的数据转化为离散的值进行处理。能够对不完整数据进行处理, 这个重要也重要, 其实也没那么重要, 缺失数据采用一些方法补上去就是了。

2.3 基于模糊神经网络的弱关联挖掘方法

(1) 数据选择

基于云计算网络环境下自主选择数据, 观察数据信息是否与弱关联挖掘算法保持一致。数据选择不仅考虑所处环境, 更要从多个角度展开研究, 观察是否存在可能影响信息结果使用效率的干扰因素。建立在安全环境下的数据选择与数据控制之间存在联系体系, 明确这一联系体系也是开展后续挖掘任务中所必须要达到的, 只有达到这一效果在信息传输中才不会受到影响。数据运行环境中所存在的各项控制问题中, 数据选择也是一个良好筛选过程, 通过建立相互联系来促进最终的数据结果稳定性程度与实际使用需求保持一致。数据选择是接下来运算任务开展的第一个基础步骤, 确保数据准确程度进入到下一阶段的挖掘计算, 数据挖掘需要结合计算机控制系统进行, 选取不同风险环境下的典型参数作为接下来的运算依据, 弱关联挖掘得到的结果才能准确反映真实情况。

(2) 数据预处理

将所选择的数据初步处理后建立起数据库, 将数据库参与挖掘算法中, 帮助节省时间, 同时促进管理效率不断提升完善。数据处理与使用需求在功能方向上保持一致, 功能确定后对信息的处理也能观察到其中是否存在影响功能进行的因素, 风险信息参数在预处理模式下便得到解决, 接下来开展各项控制管理计划也能得到充分帮助。数据挖掘处理要建立在多个控制模式下, 发现影响功能正常进行的因素后, 配合控制解决措施开展挖掘。预处理属于模糊处理, 数据库系统中可能还会存在一些多余信息, 在接下来的处理中会继续筛选, 直到信息全部为关联部分。预处理中信息干扰要排除, 并通过数据库建立来提升系统的使用安全性, 降噪、平滑处理后所得到的信息才能继续存入到数据库系统中, 为信息使用打下安全基础^[2]。

(3) 塑造训练数据

训练数据塑造需要结合信息集合进行, 将原信息特征与预处理后的信息特征进行比较, 具有相同点的部分划分为特征集合中, 用 $\{A_1, A_2, \dots, A_m\}$ 表示。集合塑造完成后对信息进行塑造训练, 用具体的数值表现出特征属性值域范围。对于可能会产生的值域使用安全问题, 建立出长期运行控制模块, 用 C 表示集合, 根据数据塑造挖掘算法来进行划分, 共分为 $C_1, C_2, \dots, C_n$, 将数据引入到塑造训练模块中, 进行综合性描述分析, 从而实现对数据运行使用效果的观察作用, 根据假设的数据集合来建立一个 $m+1$ 元组训练集合目标, 塑造描述为 $(a_1, a_2, \dots, a_m, C_i)$, 根据挖掘算法描述公式可以知 $a_j \in \text{val}(A_j) (1 \leq j \leq n)$; $C_i \in C (1 \leq i \leq n)$ 。可以将不同需求环境下的数据进行划分, 得出检测对照组与数据训练组, 两组数据之间结合对比, 所得到的结果能够准确反映出数据库运行环境, 并为所开展的各项数据体系建立同等稳定运行环境。

3 弱关联挖掘算法环境下冗余数据的产生(The generation of redundant data in the environment of weak correlation mining algorithm)

基于云计算模拟下, 信息获取渠道变得更加丰富, 信息获取数量增多后, 同时也产生了大量冗余信息, 对于这部分信息如果不能合理计算, 接下来的控制计划也会因此受到影

响。由此可见, 冗余数据产生与数据挖掘环境有直接关系, 风险评估系统的筛选甄别能力也是造成这一隐患问题的具体原因, 在当前模式下, 如果不能合理控制隐患因素, 云计算技术应用效率会因此受到影响, 冗余数据还会造成系统错误识别, 导致弱关联挖掘不能正常进行。冗余数据产生会对常规状态下的信息传输造成不良影响, 导致最终的控制能力出现问题, 建立一个解决控制方案, 对冗余数据进行划分隔离来保障有用信息的正常传输, 也是解决当前问题有效措施。研究解决控制方案要从多个角度进行, 观察开展期间的影响因素, 以及可能造成最终数据安全使用的因素, 通过这种方法来建立起长期工作环境, 从而达到最佳设计控制目标。目前解决方案研究中, 已经有技术人员提出在弱关联环境下建立一个弱小集合关联系统, 将数据挖掘定位在有效信息中, 从而必满冗余信息计算造成最终运算速度降低, 在实现技术层面上仍然需要继续深入研究。

4 建立弱聚类算法的弱关联规则挖掘模型(A weak correlation rule mining model for weak clustering algorithm is established)

4.1 数据关系匹配实现原理

实现数据之间的匹配关系, 可以从云计算环境中建立数据集合矩阵。数据匹配是自动进行的, 捕捉表达信息, 深入挖掘其中与其他数据具有关联的部分。通过建立基层数据关系来探讨最终的数据匹配关系, 对提升数据结果运行稳定性也有很大帮助。数据关系匹配需要在云计算环境下找到相同的数据参与运算, 实现数据之间相互联系, 并在联系基础上筛选出冗余部分参数, 根据产生频率来计算出具体矩阵, 观察矩阵判断不同风险隐患发生概率, 并在接下来的各项控制计划中建立起具有实际意义的工作环境。本文所研究方法与传统方法在数据挖掘数量上的比较图, 如图1所示。

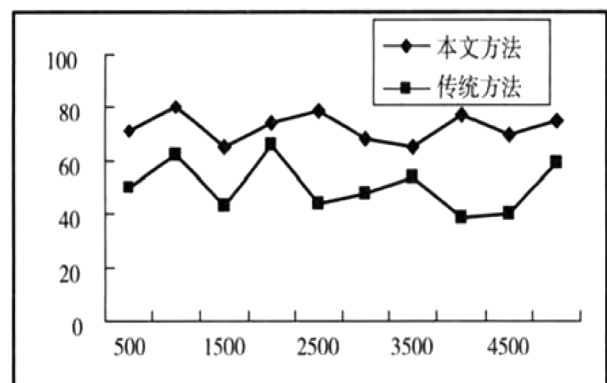


图1 弱聚类算法和传统方法挖掘数量比较

Fig.1 Weak clustering algorithm and traditional method mining quantity comparison

采用弱聚类算法是在信息挖掘数量上有明显提升, 各个节点处的信息采集数量都已经超过传统算法, 通过算法之间的相互比较可以发现, 不同模式下开展信息分析, 是帮助提升控制效果的有效方法, 使用过程中所发现的各类信息比较问题, 建立长期稳定运行环境对提升信息挖掘效率也有很大帮助。当前数据运算模式下所遇到的问题大部分是由于环境因素导致, 如果不能协调好环境问题, 数据运算模式下开展各类数据挖掘算法也会受到阻碍。

4.2 关联数据弱聚类处理

匹配任务完成后进入到数据处理阶段,通过建立基层工作联系体系来实现系统之间配合运行,对关联数据进行弱聚类处理同样基于云计算环境下开展。根据匹配所得到的数据结果信息来进行,观察到数据中存在问题时及时探讨解决方案,对于使用过程中可能会产生的问题,充分探讨解决规划方案,对提升最终任务完成效率也有很大帮助,关联模式下观察数据运行所处环境,弱聚类方法处理信息对提升关联数据也有很大的帮助,弱关联是针对少量关联模式来进行的,当前环境下所遇到的各类问题均由信息风险因素导致,采用弱聚类处理方法提升了最终结果准确程度,对于提升最终的计算效率也有很大帮助。不同类型弱聚类环境中,数据库运行所面对的风险因素有很大差异性^[3]。模拟信息处理模式需要对使用情况做出模拟,尝试在不同运行环境中所遇到的风险隐患问题,确保所提取的信息结果与实际情况之间不存在出入,弱关联环境中数据之间联系本身较少,冗余信息环境带来的阻碍影响更严重,只有解决这部分问题,接下来进行的信息对接检测才更加高效。

4.3 小区域关联划分

划分小区域关联是实现接下来挖掘算法的基础部分,划分要从多个角度进行,首先建立一个整体性的挖掘体系,观察是否具有可行性,掌握故障隐患因素后从数据关联划分角度建立一个矩阵研究体系,并观察数据运行是否处于合理的状态范围内。小区域关联信息划分完成后进入到整合阶段,按照这一划分结果对比其他方向数据库,建立起长期工作环境。大数据处理是由中多个小环境组成,因此数据划分与挖掘任务也是从各个小范围开展,冗余环境中信息有效筛选才能进入到更深层次的信息观察处理中,实现弱关联项目更深层次研究。云计算环境下信息处理速度虽然有明显提升,但在处理能力上却因此受到影响,建立这一联系体系是解决现存问题的有效方法。各个控制计划之间联系体系也需要通过方案完善不断提升运行使用效率。

4.4 弱关联环境下的挖掘过程

数据弱关联环境下,信息获取会建立独立的信道,在信道控制作用下进行其他层面数据传输,实现云计算环境下信息最稳定的获取形式。数据库中全部信息数量统计需要浪费大量时间,可以对其进行假象拟定,例如将其假设为1000,这样在运算分析中也能避免信息获取数量与实际需求不符合。信息挖掘获取需要在同步运算环境下开展,通过这种方法来帮助提升最终结果稳定程度。挖掘过程严重要首先从信息获取层面开展,统计信息数量与传输方法,挖掘过程中观察数据信息之间的关联情况,建立起具有长期控制能力的风险评估体系。信息数据传输中彼此之间具有关联性,关联性情况了解后,进入到更深层次的数据保护阶段,采用数据挖掘试探方法检验数据库使用稳定性情况,能够帮助建立起长期工作体系,并观察在系统中是否存在需要更新完善的部分。弱关联环境下数据信息之间的联系线索非常少,要确保捕捉的精准程度,信息有效利用后进入到控制层面,协调好数据之间的交流模式并建立具有保护意义的数据检索环境,结束数据挖掘检索后可以进入到更深层次的控制内容中。

5 弱关联冗余环境下的挖掘算法仿真实验(Simulation experiment of mining algorithm in weak correlation redundant environment)

确定挖掘算法开展模式后进入到仿真实验解读,检验挖掘算法是否具有可行性,建立长期工作体系来促进最终工作稳定性提升,也是解决当前矛盾冲突问题的主要原因。实验进行是对算法有效性的一次检验,观察在系统结构中是否存在可能会造成风险隐患因素的问题,将风险隔离排除。仿真实验检测弱关联冗余环境下的挖掘计算时,首先针对数据库部分的控制能力进行检验,虽然弱关联环境下数据之间联系较小,但通过这种控制挖掘方法也能帮助捕捉数据信息传输动向,确定信息之间的联系关系,从而实现对数据库运行情况的模拟检验。该方法在数据稳定性控制能力上较强,通过分析现场工作模式来建立一个长期工作环境,能够达到最佳控制效果。弱关联冗余环境是在云计算模式下展开数据挖掘的,需要在短时间内建立长期工作模式,云计算模式数据信息的数量庞大,文章方法能够帮助快速明确工作开展方法,掌握数据信息彼此之间的联系性,节省数据分类使用时间。仿真实验开展在大数据云计算环境下,实验结果显示通过建立弱关联环境中的弱聚类挖掘算法应用后效果明显,弱关联问题得到解决,并且在这一环境下所开展的各项数据运算计划也逐渐开展完善,最终进入到理想的控制环境。仿真实验分别模拟不同环境下系统运行所能遇到的问题。实验所针对的数据内容具有代表性,但并不完全表示真实运行过程,对运行情况进行分析时还需要综合考虑问题,应用该种挖掘算法开展控制效果明显,能够帮助有效提升大数据环境中的分析速度。

6 结论(Conclusion)

对于弱关联冗余环境下的挖掘算法研究,当前技术方法已经十分先进,未来发展中还要达到综合控制效果,云计算环境下提升数据处理运行效率,对最终数据处理效果提升也有很大帮助。本文提出了一种基于弱聚类算法的云计算环境下海量数据中弱关联挖掘方法,针对弱关联冗余环境进行,通过数据的描述特征对数据特征进行分解,依据数据特征,对全部数据进行融合。基于关联决策概率将云计算环境下海量数据进行有效划分,完成所有数据特征关联概率的计算,通过弱聚类方法对属性元素进行分类,将数量型元素转换成类别型,通过弱化关联规则方法,对经聚类处理后的数据进行挖掘。实验结果表明,所提方法具有很高的高效性及有效性。

参考文献(References)

- [1] 黄潮.云计算环境下的海量光纤通信故障数据挖掘算法研究[J].激光杂志,2017(1):96-100.
- [2] 张忠林,田苗凤,刘宗成.大数据环境下关联规则并行分层挖掘算法研究[J].计算机科学,2016,43(1):286-289.
- [3] 童钊,等.分布式系统中多用户网络应用的概率型调度算法研究[J].电子学报,2016,44(7):1679-1688.

作者简介:

蔡柳萍(1981-),女,硕士,讲师.研究领域:大数据,数据挖掘。