

多尺度关联规则尺度上推算算法

李超 赵书良 赵骏鹏 高琳 池云仙

(河北师范大学数学与信息科学学院 石家庄 050024)

(河北师范大学河北省计算数学与应用重点实验室 石家庄 050024)

摘要 数据挖掘在多尺度研究方面取得了一些成果。然而,多尺度数据挖掘研究还不够深入和完善。目前针对空间和图像数据的研究较多,对于一般数据的多尺度数据挖掘的研究较少。随着大数据应用的不断发展,其研究显得尤为重要。针对上述问题,进行了普适的多尺度关联规则尺度转换方法的研究。首先,基于包含度的相似理论提出频繁项集的处理方法;然后,以图像金字塔为理论基础,提出了多尺度关联规则尺度上推算算法 MSARSUA (Multi-Scale Association Rules Scaling Up Algorithm);最后,利用 H 省¹⁾ 全员人口真实数据集、UCI 公用数据集和 IBM 数据集对所提算法进行了实验验证与分析,结果表明 MSARSUA 具有较高的覆盖率、较高的 F1-measure 值和较低的平均支持度估计误差,在效率上比 Apriori 算法和 FP-Growth 算法有较大的提升,在性能上比 SU-ARMA 有更好的表现。

关键词 多尺度,关联规则,尺度上推,多尺度关联规则挖掘

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2017.08.049

Scaling-up Algorithm of Multi-scale Association Rules

LI Chao ZHAO Shu-liang ZHAO Jun-peng GAO Lin CHI Yun-xian

(College of Mathematic & Information Science, Hebei Normal University, Shijiazhuang 050024, China)

(Hebei Key Laboratory of Computational Mathematics & Applications, Hebei Normal University, Shijiazhuang 050024, China)

Abstract Great achievements have been made on multi-scale research of data mining. However, multi-scale data mining research is far from being deep and perfect. Current research, which mainly focuses on space and image data, pays less attention to multi-scale data mining on the general data. With the continuous development of big data applications, research of multi-scale data mining becomes particularly important. Regarding the issue above, this paper carried out a study of scale-conversion methods on universal multi-scale association rules data mining. First of all, this paper gave an approach of frequent items based on the similarity theory of including degree. Then, the paper proposed an algorithm named MSARSUA (Multi-Scale Association Rules Scaling Up Algorithm) based on the theory of image pyramid. Finally, experimental results on data sets from H province, UCI and IBM show that algorithm MSARSUA has higher coverage, higher F1-measure and lower estimation error of average support. Algorithm MSARSUA outperforms both Apriori algorithm and FP-Growth algorithm on efficiency aspect. Meanwhile, the results indicate that algorithm MSARSUA possesses superior performance compared with algorithm SU-ARMA.

Keywords Multi-scale, Association rules, Scaling-up, Multi-scale association rules mining

1 引言

多尺度是一种客观存在的现象,随着数学、地学、遥感学等领域研究的深入,多尺度科学已经成为一门跨学科的研究课题。国内外学者在数据挖掘领域对多尺度做了许多研究。在国内,文献[1]结合 Potts 模型和 Markov 动态过程,提出衡

量多尺度社团结构稳定性的算法,该算法可以获得社团结构在多个时间尺度的最佳数量;文献[2]基于多尺度粗糙集模型,提出一种新的决策树构造算法,有效地去除了噪声规则,提高了对噪声数据的适应能力;文献[3]将多尺度科学理论与关联规则挖掘技术相结合,提出了多尺度关联规则尺度上推算算法。在国外,文献[4]利用多尺度 AdaBoost 算法,提出多尺

¹⁾ 应用用户保密性要求,隐去省份真实名称,而称为 H 省。

到稿日期:2016-06-07 返修日期:2016-10-02 本文受国家自然科学基金项目(71271067),国家社科基金重大项目(13&ZD091),河北省高等学校科学技术研究项目(QN2014196),河北师范大学硕士基金(xj2015003)资助。

李超(1991-),男,硕士生,主要研究领域为数据挖掘、智能信息处理,E-mail:hbsd_lichao@sina.com;赵书良(1967-),男,教授,博士生导师,主要研究领域为数据挖掘、智能信息处理,E-mail:zhaoshuliang@sina.com;赵骏鹏(1990-),女,硕士生,主要研究领域为数据挖掘、智能信息处理,E-mail:zhaojunpeng115@126.com;高琳(1992-),女,硕士生,主要研究领域为数据挖掘、智能信息处理,E-mail:gaol520smile@163.com;池云仙(1987-),女,硕士生,主要研究领域为数据挖掘、智能信息处理,E-mail:lovebeyond.630@163.com。

度和分层次点集算法,对三维激光扫描点云数据进行分类,提高了分类的准确率;文献[5]提出了一种基于遥感图像多尺度分割的交互式分类方法,其分类结果好于单一尺度的分类算法,执行效率得到了显著的提升;文献[6]对原始数据进行SOM编码从而获得数据原型,采用多尺度谱聚类算法对彩色图像进行分割,减少了时间复杂度,分割效果优于Kmeans聚类方法。从目前国内外的研究情况来看,多尺度数据挖掘研究主要针对空间和图像数据,在一般数据集上不够成熟。为了扩展多尺度理论在关联规则挖掘技术中的应用,本文从以下3个方面进行研究:1)提出基于相似度的频繁项集处理方法;2)提出多尺度关联规则尺度上推算法;3)通过实验验证算法的可行性和有效性。

2 频繁项集处理方法

根据关联规则的定义,知识表现为若干条有趣的规则。关联规则挖掘的过程可分为两个步骤:1)根据设定的最小支持度阈值,找到所有的频繁项集;2)根据设定的最小置信度阈值,由频繁项集产生关联规则。由此可见,关联规则挖掘的实质是挖掘数据的频繁项集。实际应用中,可以将频繁项集作为尺度转换的对象。关联规则尺度上推算法的具体应用是在子尺度数据集上挖掘得到的频繁项集上推到父尺度,从而得到父尺度的频繁项集。

2.1 数据存储结构

为了便于收集和处子尺度数据集的频繁项集,提出数据存储结构模型。

定义1(数据存储结构) 设 U 和 A 是有限集,即 $U = \{x_1, x_2, \dots, x_n\}$, $A = \{a_1, a_2, \dots, a_n\}$ 。对于每个属性 a_i 取值有限域 V_i ($i \leq m$),且对任意 $x_i \in U, a_j \in A$,有 $f(x_i, a_j) \in V_j$ ($i \leq n, j \leq m$)。记 $F = \{f(x_i, a_j) | x_i \in U, a_j \in A\}$,则称 (U, A, F) 为数据存储结构。将频繁项集和支持度以数据存储结构的形式进行表示。例如,表1列出了频繁项集和支持度存储结构的样例。其中, U 为频繁项集集合, A 为基准尺度数据集合, V 为支持度。

表1 频繁项集与支持度存储结构

U	a_1	a_2	a_3	a_4
x_1	0.3	0	0.4	0.3
x_2	0	0.3	0.2	0.6
x_3	0.3	0	0	0.2
x_4	0.5	0.2	0.7	0.2

由于各基准尺度数据集的事务不同,因此频繁项集也不同。由表1可以看出,基准尺度数据集中不存在的频繁项集的支持度记为0,在此情况下进行尺度转换会造成频繁项集缺失。从统计学的观点来看,频繁项集作为数据集的一种统计结果,在一定程度上可以表现出数据集本身的分布与特征;统计学中的相似度是指样本集之间相似的程度,已经有学者将相似度理论引入到数据挖掘领域中来计算权重^[3]。为了减少频繁项集的缺失,提出一种基于相似度的频繁项集处理方法。根据集合之间的相似度,使用已知的频繁项集支持度估算待定的频繁项集支持度,避免了尺度转换过程中频繁项集丢失的现象,进而提高了高尺度转换的精度。

2.2 基于相似度的频繁项集处理方法

计算相似度的方法有多种,本文采用基于包含度的相似

度方法。在信息不确定、不完全的情况下,基于包含度的相似度方法可以给出与目标结果最接近的答案。下面给出包含度和相似度的定义。

定义2(包含度) 设一个 X 论域, A 和 B 是 X 中的两个子集合,集合 A 包含于集合 B 的程度 $D(B/A)$ 为包含度,若满足以下3个条件^[7]:

- (1) $0 \leq D(B/A) \leq 1$;
- (2) $A \subset B$ 时, $D(B/A) = 1$;
- (3) $A \subset B \subset C$ 时, $D(A/C) \leq D(A/B)$ 。

根据包含度的定义,给出经典集合的包含度定理和证明。

定理1 设有限集合 X ,对于 $A, B \subset X$, A 关于 B 的包含度为:

$$D(B/A) = \frac{N(A \cap B)}{N(A)} \quad (1)$$

其中, $N(A)$ 表示集合 A 中元素的个数, $N(A \cap B)$ 为集合 A 和集合 B 中相同的元素。

证明:定义2中(1)、(2)显然成立。当 $A \subset B \subset C$ 时,有 $N(A \cap B) = N(A)$, $N(A \cap C) = N(A)$,于是 $D(A/C) \leq D(A/B)$ 。因此,式(1)是有限集合 X 的幂集合 $P(X)$ 的包含度。定理1得证。

定义3(相似度) 设 $F_0(X) \subseteq F(X)$,对于任意的 $A, B \in F_0(X)$ 有数值 $S(A, B)$ 相对应,且满足条件:

- (1) $0 \leq S(A, B) \leq 1, S(A, A) = 1$;
- (2) $S(A, B) = S(B, A)$;
- (3) $A \subset B \subset C$ 时, $S(A, C) \leq S(A, B)$,称 S 为 $F_0(X)$ 上的相似度。

结合包含度理论,提出两种相似度的计算方法,采用相似度的定义对其进行证明。

定理2 设有限集合 X , D 是 $F_0(X)$ 上的包含度。对于 $A, B \subset X$, A 与 B 之间的相似度为:

$$S_1(A, B) = D(B/A) \wedge D(A/B) \quad (2)$$

$$S_2(A, B) = 0 \vee (D(B/A) + D(A/B) - 1) \quad (3)$$

证明:因为 $0 \leq D(B/A) \leq 1, D(A/A) = 1$,所以 $0 \leq S(A, B) \leq 1, S(A, A) = 1$ 。定义3中(1)、(2)显然成立。当 $A \subset B \subset C$ 时, $S_1(A, C) = D(C/A) \wedge D(A/C) = D(A/C)$, $S_1(A, B) = D(B/A) \wedge D(A/B) = D(A/B)$; $S_2(A, C) = (D(C/A) + D(A/C) - 1) \vee 0 = D(A/C)$; $S_2(A, B) = (D(B/A) + D(A/B) - 1) \vee 0 = D(A/B)$;因为 $D(A/C) \leq D(A/B)$,所以 $S_1(A, C) \leq S_1(A, B)$ 得证,即 $S_1(A, B)$ 和 $S_2(A, B)$ 均为相似度。定理2得证。

将基于包含度的相似度理论应用于处理频繁项集的方法中,通过已知频繁项集的实际支持度来估算该频繁项集的支持度。处理频繁项集的支持度的公式如下:

$$sup_i^* = \frac{1}{m} \sum_{j=1}^k S_{ij} \cdot sup_j \quad (4)$$

其中, sup_i^* 为项集 f 在集合 a_i 中的支持度估计值; sup_j 为项集 f 在集合 a_j 中的支持度的实际值; S_{ij} 为集合 a_i 和集合 a_j 之间的相似度; m 为选定频繁项集的支持度不为0的集合计数。

利用式(4)估算所有频繁项集的支持度为0的单元的值,得出频繁项集的支持度的近似值,从而完善了频繁项集-支持度的存储结构,有助于尺度上推算法的实现。

3 多尺度关联规则尺度上推算法

尺度上推旨在将小尺度信息转换为大尺度信息,通过分析样本中比较精确的信息,获取整体的一般信息,如均值、方差等。从本质上讲,尺度上推保留了数据的宏观特征,忽略了数据的微观特征。因此,尺度上推是一个“集聚”的过程。

3.1 理论基础

本文将高斯金字塔法作为尺度上推的理论基础,其本质是利用待估计点邻域内的数据点进行加权平均得到待估计点的预测值,核心在于计算加权系数,即邻域内的数据点对待估计点的贡献度。高斯金字塔方法在图像滤波中具有广泛应用,其不仅可以保证图像缩扩后的亮度平滑,而且不会出现接缝效应。高斯金字塔法包括低通滤波和下采样两个阶段。采用直接下采样方法处理图像时会出现明显的失真现象。通过高斯金字塔方法处理图像,其图像质量得到显著提高。下面详细介绍高斯金字塔的生成原理。

采用低通滤波方法处理原始图像 g_0 得到粗略图像 g_1 , 滤波是由一族局部对称的权值函数与原始图像做相同卷积的过程。设 g_0 为高斯金字塔的最底层,则其上一层 g_1 的高斯金字塔由式(5)生成。

$$G_1(i, j) = \sum_{m=-2}^2 \sum_{n=-2}^2 w(m, n) G_0(2i+m, 2j+n), 0 \leq i \leq C_L, 0 \leq j \leq R_L \quad (5)$$

其中, $w(m, n)$ 为生成核, 是一个窗口函数; C_L 和 R_L 是第 0 层的行数和列数。

由 g_0 生成 g_1 即是对 g_0 进行低通滤波后做隔行隔列的降 2 下采样操作, 算法本质上是对 g_0 求加权平均来得到 g_1 ^[8]。图 1 为高斯金字塔生成过程的一维表示, 实际上也是尺度上推的工作原理。

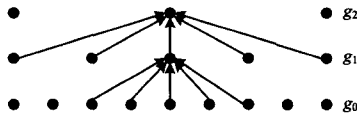


图 1 高斯金字塔的一维表示

高斯金字塔法在图像学中应用广泛且效果良好, 本文将高斯金字塔的加权平均思想推广到多尺度挖掘尺度上推算法中, 其计算公式为:

$$U(x_i)^* = \frac{\sum_{j=1}^n w_j U(x_j)}{\sum_{j=1}^n w_j} \quad (6)$$

其中, $U(x_i)^*$ 为目标尺度集合内的某个频繁项集支持度的估计值; w_j 为基准尺度集合相对于目标尺度集合的权重系数, $U(x_j)$ 为基准尺度集合内的某个频繁项集的实际支持度, n 为数据集合的个数。 w_j 的计算方法为基准尺度集合占目标尺度集合的比重, 体现了基准尺度集合在目标尺度集合上的影响程度。例如: 对于“受教育的程度”而言, 若“小学”这个子集合所占比例较大, 则“小学”的受教育程度对目标尺度集合的影响程度较大。

3.2 算法过程

MSARSUA 算法的基本思想是: 首先, 对数据进行预处理操作, 得到离散类型的数据集, 并在此基础上构建多尺度数据集; 其次, 选择基准尺度, 使用现有的关联规则算法(如

Apriori) 在基准尺度上挖掘频繁项集, 构建频繁项集与支持度存储结构, 使用相似度方法对其进行处理; 然后, 把已经完善的频繁项集与支持度存储结构作为已知样本点, 将基准尺度集合在目标尺度集合的比重作为权重系数, 加权平均得到目标尺度集合的候选项集; 最后, 依据设定的最小支持度阈值筛选出目标尺度的频繁项集, 根据设定的最小置信度阈值推导出目标尺度的关联规则。

3.3 算法描述

算法 1 MSARSUA

输入: 原始数据集, 最小支持度, 最小置信度

输出: 父尺度关联规则

1. 数据预处理, 尺度划分构建多尺度数据集, 记为 DS;
2. 选择基准尺度数据集, 记为 BS;
3. For 每个子尺度数据集 ds_{BS_i} do begin
4. 调用 Apriori 和 FP-Growth 挖掘算法, 挖掘频繁项集和支持度;
5. End For
6. 调用 $(U, A, F) = \text{BuildFrequentItemsetAndSupportStructures}(\{FI_1; Sup_1\}, \dots, \{FI_i; Sup_i\}, \dots, \{FI_n; Sup_n\})$ 函数, 构建子尺度数据集的频繁项集和支持度存储结构 (U, A, F) ;
7. For 任意两个子尺度数据集 ds_{BS_i} 和 ds_{BS_j} do begin
8. 调用相似度计算方法 $Sim_{ij} = \text{getSimilarIndex}(S, FI_i, FI_j)$; 计算两个子尺度数据集之间的相似度系数;
9. End For
10. 构建相似度矩阵, 记为 M;
11. For (U, A, F) 中值为 0 的每个单元 do begin
12. 调用频繁项集处理方法 $(U, A, F)_{opr} = \text{getEstimateStorageStructure}((U, A, F), M)$; 估算待定的频繁项集的支持度;
13. End For
14. For 每个子尺度数据集 ds_{BS_i} do begin
15. 调用计算权重方法 $\lambda_i = \text{getWeight}(ds_{BS_i}, DS)$; 得到每个子尺度数据集相对于父尺度数据集的权重;
16. End For
17. For 每个子尺度频繁项集 $f \in (U, A, F)_{opr}$ do begin
18. 调用加权平均算法 $U(x_i)^* = \frac{\sum_{j=1}^n w_j U(x_j)}{\sum_{j=1}^n w_j}$, 计算出父尺度的候选项集;
19. End For
20. 调用函数 $FI_{final} = \text{getFinalFrequentItemset}(FI_c, \text{min_sup})$; 筛选出父尺度频繁项集, 根据最小置信度生成关联规则。
21. RETURN 父尺度关联规则

4 实验分析

本文采用 UCI 数据集、IBM 数据集和 H 省全员人口数据集进行实验, 验证 MSARSUA 算法的可行性和有效性。实验环境为 CPU Intel Core i7-2600 3.40GHz、内存为 DDR3 4GBytes、操作系统为 Windows7x86 和数据库环境为 Oracle 10g。实验算法使用 Python 语言编写, 开发工具为 Python 2.7、Eclipse jdk 1.7 和 Pydv3.9。实验中采用经典 Apriori 算法挖掘基准尺度数据的频繁项集。

4.1 数据集

表 2 为实验数据集的相关信息, 包括数据集名称、项目数、事务数、平均事务长度和子孙尺度数据集个数等。其中, 项目数指数据集中不同项集的个数, 事务数指数据集中记录的条数。

表2 实验数据集的相关信息

数据集	项目数	事务数	平均事务长度	子孙尺度数据集个数
Connect	128	67557	43	15
Pumsb	2113	49046	74	15
T10I4D100K	870	80000	27	15
H省全员人口数据	36	140663	7	15

4.2 实验结果分析

本文主要从覆盖率、F1-measure、平均支持度估计误差以及运行时间 4 个方面在不同相似度、不同支持度阈值上对 MSARSUA 算法与文献[3]中的 SU-ARMA 算法进行了对比和分析,如表 3—表 6 所列。在效率方面,本文算法与 Apriori 算法和 FP-Growth 算法进行了对比,如图 2 所示。表 3—表 6

中的 S_1 代表相似度系数 $S_1(A, B)$, 如式(2)所示。

由表 3—表 6 可知,MSARSUA 的覆盖率和 F1-Measure 值较高,说明使用 MSARSUA 算法产生的假正项集和假负项集的比重较小,能够由下层尺度数据集得到上层尺度数据集的绝大部分的频繁项集和关联规则;算法的平均支持度估计误差较低,尤其在 T10I4D100K 和 Pumsb 数据集上其在 2% 以下,说明算法在估计频繁项集的支持度上表现良好;由图 2 可知,利用下层尺度数据集挖掘出来的结果推导上层尺度数据集隐含的关联规则的运行效率比直接使用 Apriori 算法或 FP-Growth 算法挖掘目标尺度数据集的关联规则的运行效率更高,在效率方面提高约 90%。

表3 MSARSUA 算法应用在 H 省全员人口数据集尺度上推部分的实验结果

评价指标	0.20		0.22		0.24		0.26		0.28	
	S_1	SU-ARMA	S_1	SU-ARMA	S_1	SU-ARMA	S_1	SU-ARMA	S_1	SU-ARMA
覆盖率	96.72	91.8	100	98.25	100	100	96.36	94.55	100	100
F1-Measure	98.33	95.73	100	99.12	99.10	99.10	98.15	97.20	100	100
平均支持度误差	3.25	3.06	2.80	2.85	2.54	2.71	2.36	2.38	2.72	3.07

表4 MSARSUA 算法应用在 Connect 数据集尺度上推部分的实验结果

评价指标	0.20		0.22		0.24		0.26		0.28	
	S_1	SU-ARMA	S_1	SU-ARMA	S_1	SU-ARMA	S_1	SU-ARMA	S_1	SU-ARMA
覆盖率	94.72	85.28	91.21	66.53	72.73	70.81	96.36	74.84	92.91	83.46
F1-Measure	92.96	89.15	92.37	77.00	83.98	82.45	89.11	82.93	91.83	88.70
平均支持度误差	5.87	8.08	6.06	5.56	5.16	5.74	4.34	5.69	3.66	4.72

表5 MSARSUA 算法应用在 T10I4D100K 数据集尺度上推部分的实验结果

评价指标	0.03		0.031		0.032		0.033		0.034	
	S_1	SU-ARMA	S_1	SU-ARMA	S_1	SU-ARMA	S_1	SU-ARMA	S_1	SU-ARMA
覆盖率	93.22	93.22	94	94	97.87	97.87	97.78	97.78	97.56	97.56
F1-Measure	94.02	94.02	94	94	98.92	98.92	98.88	98.88	97.56	97.56
平均支持度误差	1.65	1.65	1.51	1.51	1.53	1.53	1.5	1.5	1.49	1.49

表6 MSARSUA 算法应用在 Pumsb 数据集尺度上推部分的实验结果

评价指标	0.20		0.21		0.22		0.23		0.24	
	S_1	SU-ARMA	S_1	SU-ARMA	S_1	SU-ARMA	S_1	SU-ARMA	S_1	SU-ARMA
覆盖率	96.38	96.38	97.3	97.3	97.48	97.48	100	100	100	100
F1-Measure	95.73	95.73	94.43	94.43	88.89	88.89	93.92	93.92	99.38	99.38
平均支持度误差	1.87	1.87	1.69	1.69	1.63	1.63	1.5	1.5	1.55	1.55

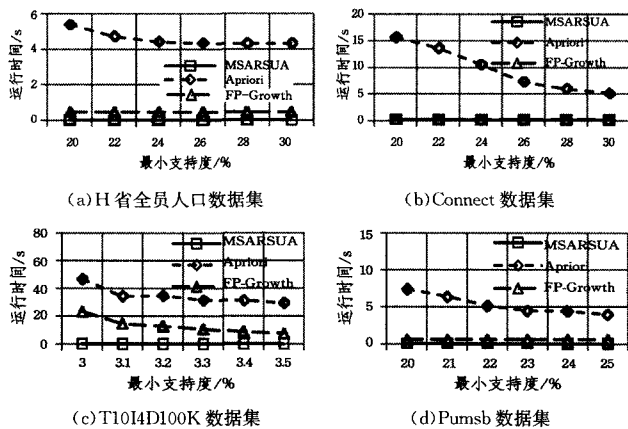


图2 MSARSUA 算法与 Apriori、FP-Growth 算法在不同数据集上运行效率的对比

将 MSARSUA 算法与文献[3]中的 SU-ARMA 算法进行对比的原因在于,MSARSUA 算法是在 SU-ARMA 算法的思想和理论上进行改进和完善的。两种算法的相同点在

于处理频繁项集的过程,都采用相似度作为处理频繁项集的权重系数。不同点在于选取的相似度方法不同;除此之外,两种算法的思想也不同,MSARSUA 算法的思想是基于高斯金字塔法的加权平均,而 SU-ARMA 算法的思想是简单的求和平均。通过对比实验发现,对于 H 省全员人口数据集和 Connect 数据集,MSARSUA 采用相似度 S_1 的表现优于文献[3]中 SU-ARMA 算法的表现,说明在处理频繁项集时,合理地选择相似度系数是十分必要的,同时也证明 MSARSUA 算法的思想优于 SU-ARMA 算法。然而,在数据集 T10I4D100K 和 Pumsb 的尺度上推过程中,相似度的选取对尺度转换的结果影响不大,所提方法的实验结果和文献[3]中 SU-ARMA 算法的结果基本相同。通过对 T10I4D100K 和 Pumsb 子孙数据集的实验研究发现,频繁项集在各子孙尺度数据集的分布均匀,由此得出基于相似度对频繁项集的处理效果基本相同,两种算法在 T10I4D100K 和 Pumsb 两个数据集上的表现都较优。此外,使用相似度系数 S_2 进行对比实验。实验结果

证明相似度系数为 S_1 的尺度上推算法的效果优于相似度系数为 S_2 的尺度上推算法。通过实验验证了 MSARSUA 算法的可行性和有效性。

结束语 本文针对多尺度数据挖掘研究尚未成熟的现状,对多尺度关联规则挖掘方法进行了分析和完善,提出了一种基于高斯金字塔法的多尺度关联规则尺度上推算法。本文基于包含度的相似理论提出了频繁项集处理方法,以便于尺度上推算法的实现;详细介绍了高斯金字塔法的加权平均思想,在此基础上提出了多尺度关联规则尺度上推算法。最后,通过实验验证了 MSARSUA 算法的可行性和有效性。未来工作中,将多尺度思想和分类、聚类等多尺度数据挖掘方法相结合,进一步研究多尺度分类和聚类挖掘算法;探索更加优秀的相似理论 and 计算权重的方法,将分布式并行的思想引入多尺度关联规则算法中,以进一步提高算法的执行效率。

参 考 文 献

- [1] LI H J, LI H Y, LI A H. Analysis of multi-scale stability in community structure[J]. Chinese Journal of Computer, 2015, 38(2): 301-312. (in Chinese)
李慧嘉,李慧颖,李爱华. 多尺度的社团结构稳定性分析[J]. 计算机学报, 2015, 38(2): 301-312.
- [2] CHEN J J, SU S B, XU H L. Decision tree optimization algorithm based on multiscale rough set model[J]. Journal of Computer Applications, 2011, 31(12): 3243-3246. (in Chinese)
陈家俊,苏守宝,徐华丽. 基于多尺度粗糙集模型的决策树优化算法[J]. 计算机应用, 2011, 31(12): 3243-3246.
- [3] LIU M M, ZHAO S L, CHEN M, et al. Scaling-up mining algorithm of multi-scale association rules mining[J]. Application Research of Computers, 2015, 32(10): 2924-2929. (in Chinese)
柳萌萌,赵书良,陈敏,等. 多尺度关联规则挖掘的尺度上推算法[J]. 计算机应用研究, 2015, 32(10): 2924-2929.
- [4] WANG Z, ZHANG L, FANG T. A Multiscale and Hierarchical Feature Extraction Method for Terrestrial Laser Scanning Point Cloud Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(5): 2409-2425.
- [5] ALEX J, HENRI P, SYLVIE, et al. Interactive Multiscale Classification of High-Resolution Remote Sensing Images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2013, 6(4): 2020-2034.
- [6] YING D, TAO G, LEI L. Self-Organizing Map Based Multi-scale Spectral Clustering for Image Segmentation[C]//2012 International Conference on Computer Science and Electronics Engineering, 2012: 329-333.
- [7] ZHANG W X, XU Z B, LIANG Y, et al. Inclusion degree theory [J]. Fuzzy Systems and Mathematics, 1996, 10(4): 1-9. (in Chinese)
张文修,徐宗本,梁怡,等. 包含度理论[J]. 模糊系统与数学, 1996, 10(4): 1-9.
- [8] YU Z M, GAO F. Laplacian pyramid and contrast pyramid based image fusion and their performance comparison[J]. Application Research of Computers, 2004, 21(10): 128-130. (in Chinese)
玉振明,高飞. 基于金字塔方法的图像融合原理及性能评价[J]. 计算机应用研究, 2004, 21(10): 128-130.
- [9] COVER T M, HART P E. Nearest neighbor pattern classification [J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.
- [10] BAI X J, ZHANG L H, LI H X. The experimental study on the intentional forgetting between two paradigms [J]. Journal of Psychological Science, 2011, 34(1): 2-6. (in Chinese)
白学军,张丽华,李红霞. 两种范式下有意遗忘的实验研究[J]. 心理科学, 2011, 34(1): 2-6.
- [11] SMITI A, ELOUEDI Z. WCOID-DG: an approach for case base maintenance based on weighting, clustering, outliers, internal detection and dbsan-gmeans [J]. Journal of Computer and System Sciences, 2014, 80: 27-38.
- [12] YAN X, FU H, TU N W. Dynamic Prediction of Coal and Gas Outburst Based on Clustering and Case-Based Reasoning [J]. Journal of Transduction Technology, 2016, 29(4): 545-551. (in Chinese)
阎馨,付华,屠乃威. 基于聚类和案例推理的煤与瓦斯突出动态预测[J]. 传感技术学报, 2016, 29(4): 545-551.
- [13] MÁNTARAS R L D, MCSHERRY D, BRIDGE D, et al. Retrieval, reuse, revision and retention in case-based reasoning [J]. The Knowledge Engineering Review, 2005, 20(3): 215-240.
- [14] SCHOOLER L J, HERTWIG R. How forgetting aids heuristic inference [J]. Psychological Review, 2005, 112: 610-628.
- [15] KENNEDY W G, TRAFTON J G. Long-term symbolic learning [J]. Cognitive Systems Research, 2007, 8: 237-247.
- [16] DERBINSKY N, LAIRD J E. Effective and efficient forgetting of learned knowledge in Soar's working and procedural memories [J]. Cognitive Systems Research, 2013, 24: 104-113.
- [17] SALAMÓ M, LÓPEZ-SÁNCHEZ M. Adaptive case-based reasoning using retention and forgetting strategies [J]. Knowledge-Based Systems, 2011, 24(2): 230-247.
- [18] COVER T M, HART P E. Nearest neighbor pattern classification [J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.
- [19] BAI X J, ZHANG L H, LI H X. The experimental study on the intentional forgetting between two paradigms [J]. Journal of Psychological Science, 2011, 34(1): 2-6. (in Chinese)
白学军,张丽华,李红霞. 两种范式下有意遗忘的实验研究[J]. 心理科学, 2011, 34(1): 2-6.
- [20] COX M T, RAM A. Introspective multistrategy learning: on the construction of learning strategies [J]. Artificial Intelligence, 1999, 112: 1-55.
- [21] ZHANG Z, YANG Q. Feature weight maintenance in case bases using introspective learning [J]. Journal of Intelligent Information Systems, 2001, 16: 95-116.
- [22] SUTTON R S, BARTO A G. Reinforcement learning: an introduction [M]. Cambridge: The MIT Press, 1998.
- [23] FRANK A, ASUNCION A. UCI machine learning repository. Irvine, CA: University of California [OL]. <http://archive.ics.uci.edu/ml>.
- [24] YANG L. Discussions of evaluation methods of cardiovascular function by pulse wave [D]. Beijing: Beijing University of Technology, 2010. (in Chinese)
杨琳. 脉搏波评价心血管功能的方法探讨 [D]. 北京: 北京工业大学, 2010.

(上接第 284 页)