[文章编号] 1007－7405(2024)04－0377－08 DOI:10.19715/j.jmuzr.2024.04.11

基于改进 Apriori 算法的高校教育满意度关联规则挖掘

陈云超[1]，谢加良[1,2]，林 玲[1,2]，刘小辉[1,3]

（1. 集美大学理学院，福建 厦门 361021；2. 数字福建大数据建模与智能计算研究所，福建 厦门 361021；
3. 厦门大学教育研究院，福建 厦门 361005）

[摘要] 针对经典关联规则 Apriori 算法在大数据集情境下易产生冗余和误导性的关联规则，以及难以确认关键性关联规则等问题，提出支持度—置信度—权重检验系数框架与后项约束的改进 Apriori 算法。首先，定义相关性系数、提升系数、错误系数并进行证明分析，进而构建权重检验系数；其次，运用主成分分析法，提取指标中的高权重影响因素作为后项，通过后项约束过滤冗余关联信息，从而筛选出更为准确的关键性关联规则。将改进的 Apriori 算法应用于高校教育满意度调查数据的关联规则挖掘并进行分析对比，实验结果验证了该算法的合理性和有效性。

[关键词] 高校教育满意度；数据挖掘；关联规则；Apriori 算法

[中图分类号] G 434；TP 311

Association Rule Mining of Higher Education Satisfaction Based on Improved Apriori Algorithm

CHEN Yunchao[1], XIE Jialiang[1,2], LIN Ling[1,2], LIU Xiaohui[1,3]

(1. School of Science, Jimei University, Xiamen 361021, China;
2. Digital Fujian Big Data Modeling and Intelligent Computing Institute, Xiamen 361021, China;
3. Institute of Educational Research, Xiamen University, Xiamen 361005, China)

Abstract: Due to the problem that the classical association rule Apriori algorithm is prone to produce redundant and misleading association rules in the context of large data sets, and it is difficult to identify key association rules, this paper proposes an improved Apriori algorithm based on the support-confidence weight-test coefficient framework and the post-term constraint. Firstly, the correlation coefficient, lifting coefficient and error coefficient were defined and proved, and then the weight test coefficient was constructed. Secondly, the principal component analysis method is used to extract the influential factors with high weight in the index as the latter term, and the redundant association information is filtered through the latter term constraints, so as to screen out more accurate key association rules. The improved Apriori algorithm is applied to mining association rules of the survey data of higher education satisfaction, and the experimental results verify the rationality and effectiveness of the algorithm.

Keywords: higher education satisfation; data mining; association rules; apriori algorithm

[收稿日期] 2022－11－27
[基金项目] 全国教育科学规划2021年度国家一般课题“高校在线教育高质量发展模式研究”（BIA210171）
[作者简介] 通信作者：林玲（1978—），博士，讲师，从事大数据技术、不确定信息决策方向研究。E-mail：35226938@qq.com

0 引言

教育质量问题成为现阶段高等教育的主要矛盾[1]，其外在可体现为社会各界对高校教育满意度的表达。教育数据挖掘（educational data mining，EDM）[2]是指将数据挖掘技术运用于教育活动实践和教育研究过程，及时发现其中存在的各种问题，以制定相关对策改善办学条件，提高教学质量，完善学习过程与教育管理。EDM 是数据挖掘与教育研究的深度融合，是教育数字化转型的必然需求[3]。

关联规则是一种重要的数据挖掘方法，其中 Apriori 算法是常用于发现频繁模式的经典关联规则算法[4-6]，被广泛运用于 EDM 中。如 Huang 等[7]将 Apriori 算法运用于分析学生成绩影响因素的挖掘，为教育管理、课程安排、学习效率和实现教学目标提供了具有较高价值的参考信息；Zhao 等[8]将 Apriori 算法运用于教学评价数据的挖掘，为提高教学质量和教育教学改革提供决策参考；Sedahmed 等[9]运用 Apriori 算法挖掘影响学生决定在高等教育机构入学的因素的重要性和相关性；郭鹏等[10]通过引入兴趣度改进 Apriori 算法并用于分析学生的成绩数据，挖掘课程之间的关联关系。Liu 等[11]利用基于兴趣的 Apriori 算法分析学生成绩，挖掘课程与课程之间的联系。黄川腾等[12]以学生学习成绩为研究对象，使用关联规则 Apriori 算法，深入探讨数据挖掘技术的实现过程，明确课程间关联关系的强弱。任鸽等[13]利用 Apriori 算法挖掘不及格课程之间的相互关联，构建基础预警规则库，在传统的支持度—置信度框架下利用提升度、兴趣度等方法筛选出强关联规则。综上，Apriori 算法在 EDM 应用中可分为两类：一是经典 Apriori 算法是对教育数据的直接应用；二是通过引入一种新的参数对所挖掘出的规则进行有效性度量。

已有的相关研究为 EDM 提供了丰富的理论指导和方法借鉴，经典 Apriori 算法在上述文献的关联规则挖掘中尚存在如下两个方面的不足：1）易产生冗余关联规则，难以得到清晰的挖掘结果；2）默认属性项的权重相同，难以提取关键性关联规则。在现实中，学校管理者需要在更少的规则里面选择关键性的规则，为高校制定相关政策以提高办学和管理水平。因此，针对上述 EDM 中经典 Apriori 算法所存在的不足，本文提出相关性系数、提升系数和错误系数构建权重检验系数，利用支持度—置信度—权重检验系数框架来删减冗余规则，通过后项约束克服默认属性项权重一致的缺点，挖掘关键性规则，即提出基于支持度—置信度—权重检验系数框架与后项约束的改进 Apriori 算法，并将改进的 Apriori 算法应用于高校教育满意度调查数据的关联规则挖掘。

1 关联规则挖掘与经典 Apriori 算法

关联规则挖掘的目标是从大数据集中找出各变量之间的关联，分析多个变量之间的联系。关联分析一般可分为：简单关联、因果关联和时序关联等三类[14]。支持度 $\mathrm{Sup}(X \rightarrow Y) = P(XY)$ 和置信度 $\mathrm{Conf}(X \rightarrow Y) = P(XY)/P(X)$ 是关联规则最核心的两个参数。支持度指的是 X 和 Y 在事务空间中同时发生的概率；置信度反映的是事务空间中 X 发生的前提下 Y 发生的概率，这两个参数的取值会直接影响最后得到的关联结果[15]。

Apriori 算法是较经典的关联规则算法，其核心是基于频繁项集性质的先验定理，算法过程可归纳为两大步骤：第一步，逐层迭代提取事务空间中所有支持度大于给定阈值的频繁项集并进行剪枝；第二步从频繁项集中获得不低于最小置信度的规则。Apriori 算法的关键是如何提取所有频繁项集，算法执行剪枝的依据是 Apriori 的先验定理。即若一个项集是频繁项集，则它的子集也一定是频繁的；若一个项集是非频繁的，则其超集也是非频繁项集。

2 改进 Apriori 算法的关联规则挖掘

2.1 经典 Apriori 算法的不足示例

当数据集较大时，传统的支持度—置信度框架易产生大量冗余规则和误导性关联[16]，从而影响正确决策。例如，有 A、B 两门课程，学生成绩分为优和差的统计结果如表 1 所示。

由表 1 可以计算得，Sup（A（优）$\rightarrow B$（优））$=0.15$，Conf（A（优）$\rightarrow B$（优））$=0.75$，表明学生在 A 课程优的前提下，B 课程优的比例为 0.75，已达到相对较高的置信度水平。但从 B 课程的角度来看，学生在 B 课程优的比例为 $0.8>0.75$，即学生在 A 课程优的情况下，反而 B 课程优的比例下降了，因此，这条规则的置信度水平虽然较高，却具有误导性。

表 1　A、B 两门课程成绩统计

Tab. 1　Statistical scores of A and B courses

项目	B(优)	B(差)	总计
A(优)	150	50	200
A(差)	650	150	800
总计	800	200	1000

针对此类问题，在获取频繁项集之后构造关联规则时，本文先定义关联规则挖掘的相关性系数、提升系数、错误系数等 3 个关联规则的检验指标；进而构建支持度—置信度—权重检验系数框架，过滤冗余规则，提高规则挖掘效率和准确度；最后通过设定关键因素的后项约束，提取关键性关联规则。

2.2　权重检验系数的构建

对数据集的关联规则挖掘过程中，需要对现实问题的多方面影响因素进行综合权衡[17]。本文通过定义相关性系数、提升系数和错误系数，进而构建权重检验系数。

定义 1　相关性系数是前项 X 和后项 Y 同时发生的概率与 XY 联合概率的差值，和 X 出现概率的比值，其计算公式为 $\mathrm{Coeff}(X\rightarrow Y)=(P(X\cup Y)-P(X)P(Y))/P(X)$。相关性系数考虑的是前项与后项的相关性，即前项是否带动后项的出现。当相关性系数为 0 时，表示 X 和 Y 的出现是相互独立的，X 和 Y 相互间不影响；当相关性系数大于 0 时，X 的出现会带动 Y 的出现；当相关性系数小于 0 时，说明 X 和 Y 是互斥的，即 X 的出现会阻碍 Y 出现。将相关性系数带入节 2.1 中的例子，$\mathrm{Coeff}(A(\text{优})\rightarrow B(\text{优}))=(0.15-0.16)/0.2=-0.05$，说明学生 A 课程优情况下，阻碍 B 课程优的出现。

定义 2　提升系数是 X 和 Y 同时发生的概率与 X 不发生而 Y 发生的概率之差，其计算公式为

$$P_r(X\rightarrow Y)=P(X\cup Y)-P(X^*\cup Y)。\tag{1}$$

其中：X^* 表示 X 不出现于事务空间，有 $P(Y)=P(X\cup Y)+P(X^*\cup Y)$，可得 $P(X\cup Y^*)=P(Y)-P(X\cup Y)$。因此，式（1）可转化为

$$P_r(X\rightarrow Y)=2P(X\cup Y)-P(Y)。\tag{2}$$

提升系数考虑的是前项对后项的影响，即前项是否出现对后项出现的影响。当提升系数为 0 时，表示 X 的是否出现对于 Y 的出现没有影响，X 和 Y 之间相互间不影响；当提升系数大于 0 时，X 的出现会带动 Y 的出现；当提升系数小于 0 时，说明 X^* 是对 Y 的出现是提升的，即 X 的不出现会带动 Y 的出现。依据式（2），$P_r(A(\text{优})\rightarrow B(\text{优}))=0.3-0.8=-0.5$，说明学生在 A 课程不是优情况下，B 课程优出现的概率会更高。

定义 3　错误系数是对 X 不出现时 Y 出现情况的度量，其计算公式为

$$\mathrm{Error}(X\rightarrow Y)=P(X^*\cup Y)/P(X^*)=\mathrm{Conf}(X^*\rightarrow Y)。\tag{3}$$

其中：有 $P(X\cup Y^*)=P(Y)-P(X\cup Y)$ 及 $P(X^*)=1-P(X)$。因此，式（3）还可表示为

$$\mathrm{Error}(X\rightarrow Y)=(P(Y)-P(X\cup Y))/(1-P(X))。\tag{4}$$

错误系数考虑的是前项不出现时，而后项出现的置信度。错误系数与置信度是对立关系，也就是说错误系数越大表明该规则越具有误导性。其取值范围为 $[0,1]$。依据式（4），$\mathrm{Error}(A(\text{优})\rightarrow B(\text{优}))=(0.8-0.15)/(1-0.2)=0.8125$，说明学生 A 课程优情况下，B 课程优的错误系数较大。

在三个检验系数中，错误系数通常数值最大，其次是提升系数，相关性系数的数值通常最小。为平衡各个系数的影响，对三个检验系数赋权而构建权重检验系数的计算公式为 $W_{\mathrm{T}}(X\rightarrow Y)=0.5\times\mathrm{Coeff}(X\rightarrow Y)+0.4\times P_r(X\rightarrow Y)-0.1\times\mathrm{Error}(X\rightarrow Y)$。在构建权重检验系数后，规定只有支持度、置信度和权重检验系数都超过相应阈值，建立支持度—置信度—权重检验系数框架。

2.3 基于后项约束的规则过滤

支持度—置信度—权重检验系数框架滤掉一些冗余的候选集，但候选集在生成关联规则过程中依然存在不具参考价值或误导性的规则。为提取关键性关联规则，本文引入高权重的评价指标作为后项约束。在关联规则挖掘过程中，权重越高的评价指标，在评价中重要性越高，而包含较高权重评价指标的后项通常也是决策者较感兴趣的规则，是问题的主要矛盾所在。

现实生活中，研究多变量的问题时，如果变量个数太多就会增加问题的复杂性。人们自然希望变量个数较少而得到的信息较多。在很多情形，变量之间具有一定的相关关系。当两个变量之间有一定相关关系时，可以解释为这两个变量携带的信息具有一定的重叠。主成分分析是一种数学统计方法，通过数学变换将一组可能存在相关性的变量转换为另一组线性不相关的变量，转换后的这组变量称为主成分。主成分分析是对于原先提出的所有变量，将关系相近的多余变量删去，建立尽可能少的新变量，实现数据降维，使得这些新变量是两两不相关的，并且新变量尽可能保持原有的信息。在实际问题中，为了全面分析问题，往往会提出具有相关的评价指标，因为每个评价指标都在不同程度上反映这个问题的某部分信息。

本文通过主成分分析法对高校教育满意度评价指标进行降维，同时根据客观数据集，提取较高权重的评价指标，参考文献[18]的方法，使计算得到的各指标权重更具客观合理性。首先，对数据集进行主成分分析，按照特征值大于 1 原则，提取 m 个主成分，并得到原始数据的初始特征值（k_1，…，k_m）、成分矩阵系数（f_{i1}，…，f_{im}）、方差的百分数（v_1，…，v_m）。然后，通过求综合评价函数确定指标权重，其中指标 i 的综合评价函数为

$$F_i = (f_{i1}/(\sqrt{k_1}\, v_1) + \cdots + f_{im}/(\sqrt{k_m}\, v_m))/(v_1 + \cdots + v_m)。$$

最后，得出各评价指标对应的权重的计算公式为

$$\omega_i = F_i/(\sum_i^n F_i), i = 1, \cdots, n。$$

由此，通过增加后项约束的方式对支持度—置信度—权重检验系数框架产生关联规则进一步过滤后，克服默认属性项权重一致，达到剔除非关键因素及误导性关联规则的目的。其中，后项约束是指后项必须包含高权重指标。

2.4 改进 Apriori 算法的关联规则挖掘

改进的 Apriori 算法进行关联规则挖掘过程如下：首先，基于目标数据集运用主成分分析法计算各指标的权重，确定后项约束；然后根据实际需要和目标数据集的特点，设定合适的最小支持度、置信度和权重检验系数；再在支持度 – 置信度 – 权重检验系数框架下产生关联规则，通过后项约束进一步过滤规则；最后，按照权重检验系数→置信度→支持度对关联规则从高到低降序排列，排序越靠前，关联规则越重要。由此，就能最大限度地挖掘出有价值的关键性关联规则。改进的 Apriori 算法关联规则生成的流程图如图 1 所示。

改进 Apriori 算法的伪代码可描述如下：

输入：事务集 D，最小支持度 min Sup，最小置信度 min Conf，最小权重检验系数 min W_T

输出：规则列表 R

1）计算指标权重，确定后项约束；

2）扫描事务集 D，产生候选 1 项集的集合 C_1，并计算其支持度；

3）依据 min Sup，产生频繁项集 L_1；

4）For（$k=2$；$L_k \neq ?$；$k++$）；

5）剪枝并生成新的候选项集 C_k，并计算 C_k 的 Sup；

6）依据 min Sup，生成频繁项集 L_k；

7）提取每一个 L_k 并计算 Conf，并比较 min Conf 后筛选；

8）提取每一个 L_k 并计算 W_T，并比较 min W_T 后筛选；

9）提取每一个 L_k 由后项约束筛选；

10）由筛选后的 L_k 生成规则列表 R；

11）输出规则列表 R。

3　实验过程与分析

本文实验数据集是厦门大学相关课题研究小组通过线上采集到的高校教育质量满意度调查数据。该问卷调查目的是了解大学生对高校教育质量的满意度情况，总共 8 个评价指标，包含课程、教学 、教师、同学、校园环境、文化、课外活动（如社团、学生会、班级工作、社会实践）、学校的管理和服务等，剔除无效记录后，数据集总共有 140 815 条完整记录。每条记录代表一名大学生对 8 个指标的评价值，对于各项指标的评价值有“十分满意”、“有些满意”、“有些不满意”和“十分不满意”四个选项，为便于计算，将其转化为数值型数据，四个选项分别对应于评价数值为：4、3、2、1，部分数据展示如表 2 所示。

表 2　高校教育质量的满意度部分评价数据

Tab. 2　Partial evaluation data of satisfaction degree of higher education quality

序号	课程	教学	教师	同学	校园环境	校园文化	课外活动	管理和服务
1	4	3	4	4	3	4	4	3
2	2	4	3	3	1	3	3	2
3	3	4	3	3	4	3	3	3
4	3	2	3	3	3	3	3	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

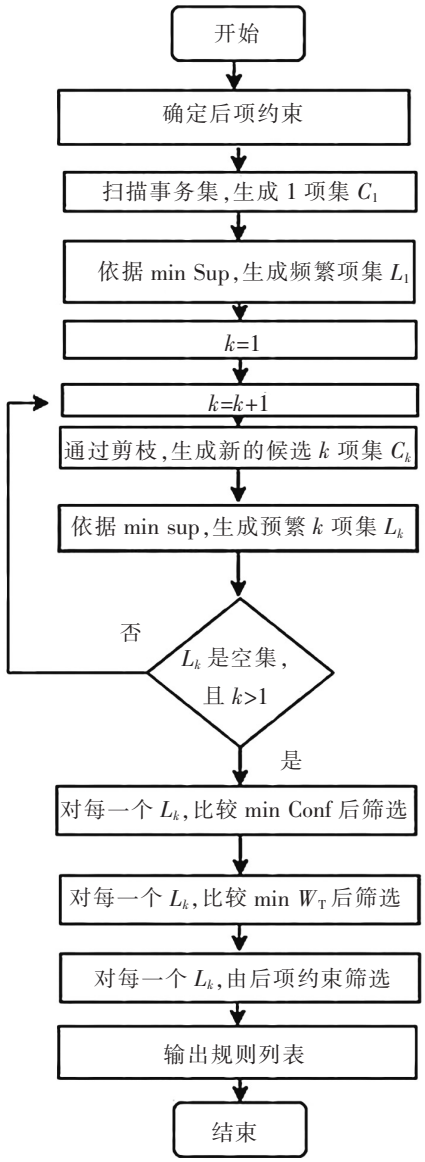

图 1　改进 Apriori 算法流程图

Fig.1　Flowchart of improving Apriori algorithm

运用社会科学统计软件包 SPSS 22.0 进行主成分分析。先将所得的数据进行 KMO 和 Bartlett 球形度检验，得到 KMO 样本值为 0.863，卡方值为 1 240.772，表明该数据集适合采用主成分分析法。主成分分析得到解释总方差如表 3 所示，成分矩阵如表 4 所示。

由表 3 可知，特征值大于 1 的主成分的累计方差值为 67.714%，在以往的文献中，达到可信的程度，原有的 8 个评价指标可以降维成主成分 1 和主成分 2。

由表 4 可知，同学、校园环境、校园文化、课外活动、管理和服务等评价指标为主成分 1，课程、教师、教学等评价指标为主成分 2。

按照节 2.2 和节 2.3 的方法计算各评价指标的权重。以“课程”指标为例，其权重为：F(课程) = ((0.74 × 53.54/$\sqrt{4.307}$) + (0.443 × 13.874/$\sqrt{1.11}$))/(53.54 + 13.874) = 0.3697。将各评价指标权重归一化处理，得到归一化的“课程”指标权重：F(课程)/(F(课程) + F(教师) + F(教学) + F(同学) + F(校园环境) + F(校园文化) + F(课外活动) + F(管理和服务)) = 0.164。用同样的方法计算其他各项指标权重，结果如表 5 所示。

表 3 解释总方差

Tab. 3 Elucidates the total variance

成分	初始特征值			提取载荷平方和		
	总计	方差/%	累积/%	总计	方差/%	累积/%
1	4. 307	53. 840	53. 840	4. 307	53. 840	53. 840
2	1. 110	13. 874	67. 714	1. 110	13. 874	67. 714
3	0. 689	8. 615	76. 328			
4	0. 539	6. 743	83. 071			
5	0. 435	5. 444	88. 515			
6	0. 395	4. 932	93. 447			
7	0. 299	3. 733	97. 180			
8	0. 226	2. 820	100. 000			

表 4 成分矩阵

Tab. 4 Component matrix

指标	成分 1	成分 2
课程	0. 740	0. 443
教师	0. 703	0. 504
教学	0. 776	0. 465
同学	0. 618	−0. 080
校园环境	0. 743	−0. 285
校园文化	0. 807	−0. 337
课外活动	0. 698	−0. 350
学校的管理和服务	0. 768	−0. 346

表 5 指标权重归一化结果

Tab. 5 The result of the normalization of indicator rights

指标	课程	教师	教学	同学	校园环境	校园文化	课外活动	管理和服务
权重	0. 164	0. 166	0. 172	0. 098	0. 101	0. 109	0. 090	0. 100

由表 5 可知教师、教学和课程等指标权重较高，其权重之和就为 0. 502，说明学校管理者应对这几个指标予以重视。根据实际情况，可选取后项含有教师、教学和课程的关联规则（在实际的应用中，可根据需要，调整纳入相对权重较高的指标）。经过多次调整，最终将最小支持度设为 0. 2，最小置信度设为 0. 6，权重检验系数设置为 0. 1，运用改进的 Apriori 算法对数据集挖掘得到 5 条关联规则，如表 6 所示。

表 6 支持度—置信度—权重检验系数框架关联规则

Tab. 6 Supportability-confidence-weight test coefficient frame association rules

序号	规则	支持度	置信度	W_T
1	教师:3→课程:3	0. 43	0. 76	0. 210 6
2	课程:3→教师:3	0. 43	0. 78	0. 200 4
3	教学:4→教师:4	0. 38	0. 89	0. 194 7
4	教师:3→教学:3	0. 32	0. 62	0. 154 8
5	教学:3→教师:3	0. 36	0. 76	0. 103 8

分析上述提取到的关联规则，从表 6 中的规则 1、2，可以发现教师与课程是相互促进的关系，学生对教师“有些满意”通常意味着对课程也“有些满意”。规则 4 表明了，学生对教师“有些满意”也会对教学“有些满意”，但从其得分看，学生对学校的教学的满意度并不太高，未达到形如规则 3“教学：4→教师：4”的强规则。规则 3、5 表明，学生如果学校教学感到满意，就意味着他们对教师比较满意，因此，教育管理者应该重视教师业务能力水平的发展，提高课堂教学质量，使得教师与学生达成良性互动，形成教与学正反馈。

为进行对比，运用支持度—置信度框架下的经典 Apriori 算法挖掘得到 32 条关联规则，并计算该部分关联规则的提升度[13]与兴趣度[17]，其中提升度公式为 $\mathrm{Lift}(X\rightarrow Y) = P(X\cup Y)/(P(X)P(Y))$，兴趣度公式为 $\mathrm{Inte}(X\rightarrow Y) = (\mathrm{Conf}(X\rightarrow Y) - P(Y))/\max(\mathrm{Conf}(X\rightarrow Y), P(Y))$，部分规则展示如表 7 所示。结果显示，在经典的支持度—置信度框架下，所筛选出来的关联规则数 32 条，6 倍于运用改进的在支持度—置信度—权重检验系数框架挖掘得到的关联规则数，造成决策者难以直接得出挖掘结论。设最小提升度为 1. 3，在支持度—置信度—提升度框架下，得到了 25 条规则，也难以得出

结论。设最小兴趣度为 0.3，在支持度—置信度—兴趣度框架下，得到了 10 条规则，其中规则“校园文化：4→校园环境：4”、“学校的管理和服务：3→课外学习活动：3”等并不是决策者所关心的关键规则。即这三种框架都没有考虑到规则的权重检验系数和后项约束，得到过多的冗余规则，无法突出关键性关联规则。

表 7　支持度—置信度框架关联规则展示

Tab. 7　Supportability-confidence framework association rules display

序号	规则	支持度	置信度	提升度	兴趣度
1	教学:4→教师:4	0.22	0.89	2.27	0.56
2	校园文化:4→校园环境:4	0.34	0.84	1.89	0.48
3	课程:3→教师:3	0.43	0.78	1.37	0.27
4	教师:3→课程:3	0.43	0.76	1.37	0.30
5	教学:3→教师:3	0.36	0.76	1.32	0.25
6	校园文化:3→教师:3	0.33	0.73	1.27	0.22
7	校园环境:3→校园文化:3	0.29	0.71	1.57	0.37
8	学校的管理和服务:3→课外学习活动:3	0.33	0.70	1.34	0.26
⋮	⋮	⋮	⋮	⋮	⋮

为验证改进算法的效率，分别采用传统支持度—置信度、支持度—置信度—支持度 、支持度—置信度—兴趣度框架与本文提出的算法效率进行对比。选用最小支持度为 0.2，最小置信度阈值为 0.6，最小提升度为 1.3，最小兴趣度为 0.3。结果如图 2 所示。

从图 2 可以看出，数据集指标项数为 8，记录集为 14 万多条的情况下，传统的支持度—置信度框架所挖掘出来的关联规则明显较多，而兴趣度框架得到的规则数相对较少，随着置信度的降低，所得到的关联规则数迅速增大。到置信度为 0.5 的时候，本文提出的方法相对于其他框架，已经产生了相对明显优势。显然，运用本文改进 Apriori 算法所提取的规则明显更为精简，且随着置信度值的降低其过滤冗余信息效果尤为明显，充分表明了在设置权重检验系数和设置适当后项关键因素约束的情况下，改进的 Apriori 算法能够较大程度地过滤冗余和误导性规则，提高了关联规则挖掘的准确度。

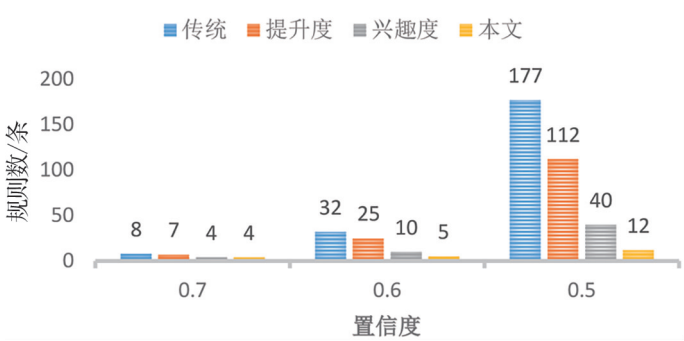

图 2　两种方法效果的对比

Fig.2　Comparison of the effects of the two methods

4　结语

教育数字化转型背景下，教育实践与研究过程中不断产生海量相关数据，高效、准确地对教育大数据进行挖掘的关键是运用合适的数据挖掘算法。本文针对经典关联规则 Apriori 算法在大数据集情境下易产生的冗余规则和误导性关联规则，与默认项之间权重是一致的问题，提出了支持度—置信度—权重检验系数框架与后项约束的改进 Apriori 算法。定义了相关性系数、提升系数、错误系数，并进行合理性分析证明，继而构建权重检验系数。运用主成分分析法，提取指标中高权重的影响因素，根据现实情形，选取关键性指标作为后项；通过后项约束过滤冗余信息，从而快速筛选出更为准确的

关键关联规则。将改进的 Aprior 算法应用于高校教育满意度调查数据的关联规则挖掘，通过与其他框架进行实验对比分析，验证了该算法的合理性和有效性，并能得到较为准确的关联规则，为高校提高办学和管理水平提供决策参考。本文改进的 Apriori 算法具备较大的灵活性，对于不同的数据集，可根据现实情境调整权重检验系数，再由主成分分析的结果，灵活确定若干关键影响因素，从而设置后项约束，过滤冗余和误导性关联规则，使得挖掘结果更为高效准确。

[参 考 文 献]

[1]王智超,朱太龙. 高等教育高质量发展的价值逻辑探寻[J]. 中国电化教育,2021(9):1-8,17.

[2]于方,刘延申. "以用户为中心"的教育数据挖掘应用研究[J]. 电化教育研究,2018,39(11):69-77.

[3]胡姣,彭红超,祝智庭. 教育数字化转型的现实困境与突破路径[J]. 现代远程教育研究,2022,34(5):72-81.

[4] DJENOURI Y,DJENOURI D,HABBAS I,et al. How to exploit high performance computing in population-based metaheuristics for solving association rule mining problem[J]. Distributed and Parallel Databases,2018,36(2):369-397.

[5]JAVED M F,NAWAZ W,KHAN K U. HOVA-FPPM:flexible periodic pattern mining in time series databases using hashed occurrence vectors and apriori approach[J]. Scientific Programming,2021(1):1-14.

[6]DHINAKARAN D,PRATHAP P M J. Protection of data privacy from vulnerability using two-fish technique with Apriori algorithm in data mining[J]. The Journal of Supercomputing,2022,78(16):17559-17593.

[7]HUANG H, TANG M J, ZENG Q T, et al. Application of student achievement analysis based on Apriori algorithm[C]//International Conference on Information Technology and Computer Application (ITCA). Guangzhou: IEEE, 2020:19-22. DOI: 10.1109/ITCA52113.2020.00011.

[8]ZHAO Y. Research on the application of university teaching management evaluation system based on Apriori algorithm[C]// 2021 2nd International Conference on Computer Information and Big Data Applications. Wuhan: IOP Publishing, 2021, 1883 (1): 012033. DOI:10.1088/1742-6596/1883/1/012033.

[9]SEDAHMED Z M,NOURELDIEN N A. Factors influencing students decisions to enrollment in sudanese higher education institutions[J]. Intelligent Information Management,2019,11(4):61-76.

[10]郭鹏,蔡骋. 基于聚类和关联算法的学生成绩挖掘与分析[J]. 计算机工程与应用,2019,55(17):169-179.

[11]LIU Y X, ZHOU F, XIN X Y. Student performance mining based on kernel density estimation interval and association rules [C]//2021 3rd International Conference on Computer Science and Technologies in Education(CSTE). New York: IEEE, 2021:58-63.

[12]黄川腾,蒲爽,唐迪,等. 基于关联规则挖掘算法 Apriori 的土木工程课程相关性分析[J]. 中国教育信息化,2020, 482(23):55-58,84.

[13]任鸽,吴猛,汗古丽·力提甫,等. 基于改进 Apriori 算法的高校课程预警规则库构建[J]. 计算机系统应用,2021,30(7):290-295.

[14]CHENG Y, YING X. Research and improvement of Apriori algorithm for association rules[C]//2010 2nd International Workshop on Intelligent Systems and Applications. Wuhan:IEEE,2010:1-4.

[15]WANG H B,GAO Y J. Research on parallelization of Apriori algorithm in association rule mining[J]. Procedia Computer Science,2021,183:641-647.

[16]GUO X Q R. Algorithm of non-redundant association rules based on user interest orientation[J]. International Journal of Performability Engineering,2018,14(8): 1745.

[17]王桌芳,赵会军,李聪,等. 基于兴趣度度量的多类差异数据关联规则挖掘[J]. 计算机应用与软件,2019,36(12):60-65,105.

[18]韩小孩,张耀辉,孙福军,等. 基于主成分分析的指标权重确定方法[J]. 兵工装备学报,2012,33(10):124-126.

(责任编辑　彭海滨　英文审校　黄振坤)