Advances in Applied Mathematics 应用数学进展, 2023, 12(11), 4898-4905
Published Online November 2023 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2023.1211482

Hans汉斯

关联规则挖掘中闭频繁项集的理解与探索

万　鑫[*]，张慧娜[*]，李裕梅，王　鑫

北京工商大学数学与统计学院，北京

收稿日期：2023年10月28日；录用日期：2023年11月23日；发布日期：2023年11月29日

摘　要

关联规则挖掘，通过数据挖掘事务之间的关联关系，被广泛应用到各个领域，主要是通过频繁项集产生关联规则，而频繁项集的挖掘又归结到闭频繁项集的挖掘，由此可见闭频繁项集在关联规则挖掘种的重要作用。本文以购物篮关联规则分析为场景，对闭频繁项集的理论进行了梳理，针对闭包算子定义中的函数i和t进行了单调性证明；对闭包算子定义满足的三条性质进行了证明；对频繁项集和其闭包的支持度进行了探讨；对封闭频繁项集和其超集之间的支持度进行了关系讨论。最后，按照频繁项集及其闭包所形成的等价类的性质进行了研究，给出了有关定理和证明，以及结论等。

关键词

关联规则挖掘，频繁项集，支持度，闭包，闭频繁项集

Understanding and Exploration of Closed Frequent Itemsets in Association Rule Mining

Xin Wan[*], Huina Zhang[*], Yumei Li, Xin Wang

School of Mathematics and Statistics, Beijing Technology and Business University, Beijing

Received: Oct. 28th, 2023; accepted: Nov. 23rd, 2023; published: Nov. 29th, 2023

Abstract

Association rule mining, is widely used in various fields by mining the association relationships between transactions through data mining, in which, association rules are mainly generated through frequent itemsets, and the mining of frequent itemsets is reduced to the mining of closed

[*]共同第一作者。

文章引用: 万鑫, 张慧娜, 李裕梅, 王鑫. 关联规则挖掘中闭频繁项集的理解与探索[J]. 应用数学进展, 2023, 12(11): 4898-4905. DOI: 10.12677/aam.2023.1211482

frequent itemsets, and this shows the important role of closed frequent itemsets in association rule mining. This article takes the analysis of shopping basket association rules as a scenario to sort out the theory of closed frequent itemsets, and proves the monotonicity of functions *i* and *t* in the definition of closure operators; proves three properties that the closure operator definition satisfies; explores the support of frequent itemsets and their closures; discusses the relationship between the support of closed frequent itemsets and their supersets. Finally, the properties of the equivalence class formed by the frequent itemsets and their closure are studied, and relevant theorems, proofs, and conclusions are provided.

Keywords

Association Rule Mining, Frequent Itemsets, Support, Closure, Close Frequent Itemsets

1. 引言

频繁项集挖掘最早是由 Agrawal 等[1]于 1994 年在挖掘关联规则的 Apriori 算法中提出来的，该算法对顾客的购物篮数据进行分析，进而通过频繁项集产生强关联规则，分析顾客在购物篮中放置的不同物品之间的联系来分析顾客的购买习惯。然后，关联规则挖掘被应用到各个领域，比如：蛋白质 DNA 结合序列模式的发现[2]，教育中的学生管理[3]，临床事件的提取和可视化[4]，动物迁徙的优化[5]等。

对大型事务数据集进行频繁项集挖掘的一个主要挑战是：满足最小支持度阈值的频繁项集往往数量较大，尤其是最小支持度阈值的设置较低时。这是因为当一个项集是频繁项集时，其包含的子项集也都是频繁项集，而大的频繁项集往往包含指数规模的子频繁项集。为解决这一问题，学者们提出了闭频繁项集和最大频繁项集的挖掘算法[6] [7]。闭频繁项集与最大频繁项集提供了完全频繁项集的紧凑表示，同时闭频繁项集还提供了项集的支持度信息[8] [9]，即一个项集和它的闭项集有相同的支持度[10]。这样，挖掘关联规则的第一步就转变为找出所有闭频繁项集。挖掘闭频繁项集也主要分为两步：(1) 找出频繁项集；(2) 检查频繁项集是否是闭集。

挖掘闭频繁项集的算法方面进展比较多[11] [12] [13] [14]，闭频繁项集挖掘的应用也比较广泛[15] [16]，但对于闭频繁项集本身理论的研究比较少。我们对其理论进行了仔细梳理，发现还有很多细节值得深思和理解。本文主要以顾客购买商品的场景来理解闭频繁项集的概念、性质及相关结论等，给出相应的定理及证明，对闭频繁项集的理论框架有更深入的剖析，并配合实例进行了演示，希望对闭频繁项集的理论发展起到比较深入的促进作用。

2. 闭频繁项集相关定义、定理

令 $I=\{x_1,x_2,\cdots,x_m\}$ 为一组称作项(item)的元素的集合。集合 $X\subseteq I$ 称为项集(itemset)。I 可以表示超市中售卖的所有商品的集合、一个网站所有页面的集合等。一个基数大小为 k 的项集称为 k 项集。我们用 $I^{(k)}$ 表示所有 k 项集的集合，即所有大小为 k 的、I 的子集[10]。

令 $T=\{t_1,t_2,\cdots,t_n\}$ 为另一个由所谓的事务标识符(tid)构成的集合[10]。集合 $\ddot{T}\subseteq T$ 称为一个事务标识符集。我们假设项集和事务标识符集都是按照字母顺序排列的。T 可以表示购买商品的顾客集合。

一个二元数据库 D 表示了事务标识符集合和项集之间的二元关系[10]，即 $D\subseteq T\times I$。我们说事务标

识符 $t \in T$ 包含项 $x \in X$，当且仅当 $(t,x) \in D$。这里的 $(t,x) \in D$，在商品购买过程中，可以表示顾客 t 购买了商品 x。

下面，为了给出闭频繁项集的定义，先给出两个相关函数的定义[10] [17]：

定义 1. 函数 $t: 2^I \to 2^T$，定义为 $t(X) = \{t \in T \mid t \text{ 包含 } x\}$，其中 $X \subseteq I$，且 $t(X)$是包含项集 X 中所有项的事务标识符集合。

定义 2. 函数 $i: 2^T \to 2^I$，定义为 $i(\ddot{T}) = \{x \mid \forall t \in \ddot{T}, t \text{ 包含 } x\}$，其中 $\ddot{T} \subseteq T$，且 $i(T)$是事务标识符集 T 中所有事务的公共项集的集合。具体地说，$i(t)$是事务标识符 $t \in T$ 中包含的项的集合。

例 1 [10]. 商品 A、B、C、D、E 被 6 个顾客购买的情况如下，第一个表格是 6 个顾客购买情况的二进制表示，其中 0 表示没购买、1 表示购买了某个商品；第二个表格是 6 个事务标识符对应商品情况，即 6 个顾客的事务标识符是 $t_1, t_2, \cdots, t_6$，其中

	A	B	C	D	E
1	1	1	0	1	1
2	0	1	1	0	1
3	1	1	0	1	1
4	1	1	1	0	1
5	1	1	1	1	1
6	0	1	1	1	0

t	$i(t)$
t_1	ABDE
t_2	BCE
t_3	ABDE
t_4	ABCE
t_5	ABCDE
t_6	BCD

每位顾客的购买情况分别表示为

$$i(t_1) = \{\text{ABDE}\}, i(t_2) = \{\text{BCE}\}, \cdots, i(t_6) = \{\text{BCD}\}$$

顾客们购买物品的交叉情况如

$$i(t_1,t_2) = \{\text{BE}\}, i(t_1,t_2,t_3) = \{\text{BE}\}, i\{t_1,t_2,t_3,t_4,t_5\} = \{\text{BE}\}, i(t_1,t_2,t_3,t_4,t_5,t_6) = \{\text{B}\}.$$

单件商品被顾客购买的情况：

$$t(\text{A}) = \{t_1 t_3, t_4, t_5\}, t(\text{B}) = \{t_1,t_2,t_3,t_4,t_5,t_6\}, t(\text{C}) = \{t_2,t_4,t_5,t_6\},$$
$$t(\text{D}) = \{t_1,t_3,t_5,t_6\}, t(\text{E}) = \{t_1,t_2,t_3,t_4,t_5\};$$

多件商品共同一起被顾客购买的情况：

$$t(\text{AB}) = \{t_1,t_3,t_4,t_5\}, t(\text{AC}) = \{t_4,t_5\}, \cdots, t(\text{ABC}) = \{t_4,t_5\}, \cdots, t(\text{ABCDE}) = \{t_5\}.$$

这里讨论一下函数 t 和 i 的单调性，在文献[17]中提出来这样的性质，但并没有给出相关证明，我们针对顾客购买商品的情况下进行了如下证明。

定理 1. 如果 $X_k \subseteq X_q$，则 $t(X_q) \subseteq t(X_k)$，即 t 是单调减函数；如果 $T_k \subseteq T_q$，则 $i(T_q) \subseteq i(T_k)$，即 i 是单调减函数。

证明：(1) $t(X_k)$ 和 $t(X_q)$ 分别表示共同购买了商品集合 X_k 和 X_q 的事务集合，而 $X_k \subseteq X_q$，则共同购买了 X_q 的顾客不会比共同购买的 X_k 多，因为购买了 X_q 的顾客必然购买了 X_k，所以有 $t(X_q) \subseteq t(X_k)$。说明函数 t 是单调减的。

(2) $i(T_k)$ 表示事务集 T_k 里共同购买过的商品集合，$i(T_q)$ 表示事务集 T_q 里共同购买过的商品集合，而 $T_k \subseteq T_q$，说明 T_q 集合里除了有 T_k 的事务集之外还有更多的事务标识符，则能从 T_q 里找到的共同商品

集合不会多于从 T_k 里找到的共同商品集合，所以有 $i\left(T_q\right)\subseteq i\left(T_k\right)$，即 i 函数的单调减的。

关于上面单调性定理，举例如下面例 2。

例 2. 根据例 1 的商品购买情况，我们假设 $X_1=\{B\}$，$X_2=\{AB\}$，则有 $t\left(X_1\right)=t\left(\mathrm{B}\right)=\left\{t_1,t_2,t_3,t_4,t_5,t_6\right\}$，而 $t\left(X_2\right)=t\left(\mathrm{AB}\right)=\left\{t_1,t_3,t_4,t_5\right\}$，可见 $t\left(\mathrm{AB}\right)\subseteq t\left(\mathrm{B}\right)$；假设 $i\left\{t_1,t_2,t_3\right\}=\left\{\mathrm{BE}\right\}$，而 $i\left\{t_1,t_2,t_3,t_4,t_5,t_6\right\}=\left\{\mathrm{B}\right\}$，有 $i\left\{t_1,t_2,t_3,t_4,t_5,t_6\right\}\subseteq i\left\{t_1,t_2,t_3\right\}$。

为了方便起见，没有严格使用集合的表示，例如：我们用 $i(t)$表示 $i(\{t\})$，用 $i(t_1t_2)$表示 $i(\{t_1,t_2\})$，用 $t(\mathrm{A})$表示 $t(\{\mathrm{A}\})$，用 $t(\mathrm{AB})$表示 $t(\{\mathrm{A,B}\})$等。

定义 3. 项集 X 的支持度为[10]：$\sup\left(X\right)=\left|\left\{t\mid\left(t,x\right)\in D\right\}\right|$，也就是商品集合 X 被购买的次数，或者说是商品集合 X 被购买的顾客数。

定义 4. 闭包算子(closure operator)为 c：$2^I\rightarrow2^I$，定义如下[10]：

$$c\left(X\right)=i\circ t\left(X\right)=i\left(t\left(X\right)\right),$$

且 $c(X)$称为 X 的闭包。

关于闭包算子的举例如下面例 3。

例 3. 根据例 1 求几个项集的闭包，

(1) $c\left(\mathrm{AB}\right)=i\circ t\left(\mathrm{AB}\right)=i\left(t\left(\mathrm{AB}\right)\right)$，这里 $t\left(\mathrm{AB}\right)=\left\{t_1,t_3,t_4,t_5\right\}$，所以

$$i\left(t\left(\mathrm{AB}\right)\right)=i\left(\{t_1,t_3,t_4,t_5\}\right)=i\left(t_1,t_3,t_4,t_5\right)=\left\{\mathrm{ABE}\right\},$$

于是，{ABE}是{AB}的闭包。

(2) $c(\mathrm{CD})=i\circ t\left(\mathrm{CD}\right)=i\left(t\left(\mathrm{CD}\right)\right)$，这里 $t\left(\mathrm{CD}\right)=\left\{t_5,t_6\right\}$，所以

$$i\left(t\left(\mathrm{CD}\right)\right)=i\left(\left\{t_5,t_6\right\}\right)=i\left(t_5,t_6\right)=\left\{\mathrm{BCD}\right\},$$

于是，{BCD}称为{CD}的闭包。

(3) $c\left(\mathrm{BD}\right)=i\circ t\left(\mathrm{BD}\right)=i\left(t\left(\mathrm{BD}\right)\right)$，这里 $t\left(\mathrm{BD}\right)=\left\{t_1,t_3,t_5,t_6\right\}$，所以

$$i\left(t\left(\mathrm{BD}\right)\right)=i\left(\left\{t_1,t_3,t_5,t_6\right\}\right)=i\left(t_1,t_3,t_5,t_6\right)=\left\{\mathrm{BD}\right\},$$

于是，{BD}是自己的闭包。

(4) $c\left(\mathrm{ABD}\right)=i\circ t\left(\mathrm{ABD}\right)=i\left(t\left(\mathrm{ABD}\right)\right)$，这里 $t\left(\mathrm{ABD}\right)=\left\{t_1,t_3,t_5\right\}$，所以

$$i\left(t\left(\mathrm{ABD}\right)\right)=i\left(t_1,t_3,t_5\right)=\left\{\mathrm{ABDE}\right\}.$$

于是，{ABDE}是{ABD}的闭包。

理解 1： 闭包，就是定义 2 和 3 中 t 函数和 i 函数复合作用的结果，其过程相当于先根据商品找到它们的共同购买者，然后再看这些共同购买者购买的项数最多的共同商品集合，也就是说，闭包运算是一个从商品集合到商品集合的一个映射过程。

定理 2. $c(X)$是 X 的闭包，则 $\sup\left(c\left(X\right)\right)=\sup\left(X\right)$。

证明：$c\left(X\right)=i\left(t\left(X\right)\right)$，则 $\sup\left(c\left(X\right)\right)=\sup\left(i\left(t\left(X\right)\right)\right)$，分解理解 1，$t(X)$是购买商品集合 X 的顾客集合，而 $i(t(X))$是顾客集合 $t(X)$共同购买的最大商品集合，寻找闭包的这个过程中，顾客集合 $t(X)$没有改变，这群顾客既购买了 X 商品集合，也是购买了 $i(t(X))$这个商品集合，所以 $\sup\left(i\left(t\left(X\right)\right)\right)=\sup\left(X\right)$，所以 $\sup\left(c\left(X\right)\right)=\sup\left(X\right)$。

关于闭包算子的如下定理3的性质，格论的文献里给的闭包算子本身就用这3条特性进行定义的[18]，当然就没有给出相关证明。我们在这里沿着文献[17]给出的闭包算子的定义，然后证明该算子在商品购买的情景中满足这三条特性。文献[10]里也有这三条性质，但没有给出证明，我们这里利用前面的定理 1 和定理 2 来给出这三条性质的证明。

定理 3. 闭包算子 c 将项集映射到项集，且满足如下 3 个特性[10]：

(1) 递增性(extensive)： $X \subseteq c(X)$；

(2) 单调性(monotonic)：若 $X_i \subseteq X_j$，则 $c(X_i) \subseteq c(X_j)$；

(3) 幂等性(idempotent)： $c(c(X)) = c(X)$。

证明：

(1) 假设从商品集 X_1 出发，共同购买了商品集 X_1 的事务集合有 $t_{i1}, t_{i2}, \cdots, t_{ik}$，则 $t_{i1}, t_{i2}, \cdots, t_{ik}$ 至少共同购买了商品集 X_1，或者还共同购买了除 X_1 之外的其他商品，所以 $X \subseteq c(X)$。

(2) 设 $X_i \subseteq X_j$，根据定理 1，有 $t(X_j) \subseteq t(X_i)$，表明事务集 $t(X_j)$ 更小且又包含在事务集 $t(X_i)$ 中，再根据定理 1，就有 $i(t(X_i)) \subseteq i(t(X_j))$，从而得到

$$c(X_i) = i \circ t(X_i) = i(t(X_i)) \subseteq i(t(X_j)) = i \circ t(X_j) = c(X_j).$$

(3) $t(X)$是寻找购买过 X 的顾客，$t(c(X))$是共同购买过 $c(X)$的顾客，而 $c(X)$是 X 的闭包、二者具有相同的顾客群，所以 $t(c(X)) = t(X)$，所以 $i(t(c(X))) = i(t(X))$。

定义 5 [10]. 若项集 X 满足 $c(X) = X$，则称其为封闭的(closed)，即 X 是闭包算子 c 的一个定点，即 X 被称为封闭项集。另一方面，若 $c(X) \neq X$，则 X 是不封闭的，集合 $c(X)$称为 X 的闭包。

理解 2： 每一个频繁项集 X，其闭包一定存在。

这是因为，根据理解 1，对于一个被购买频繁的商品集合 X，一定能找到共同购买过 X 的顾客集合 T，再反过来查找这群顾客集合 T 共同购买过的最大商品集合 Z，那么这个商品集合 Z 就是 X 的闭包，这个过程一定能做到，所以 X 的闭包一定存在。

关于一个商品集合是否封闭，举例如下面例 4。

例 4. 比如上面例 3(1)中 $c(\mathrm{AB}) = \{\mathrm{ABE}\} \neq \{\mathrm{AB}\}$，这说明{ABE}是{AB}的闭包；而例 2(3)中 c(BD) = {BD}，这说明{BD}是封闭的。

根据闭包算子的性质，可以知道 X 和 $c(X)$有相同的事务标识符集. 由此可知，若频繁集 $X \in F$ (F 是所有频繁项集的集合)没有与之频率相同的频繁超集，则它是封闭的；因为根据定义，它是事务标识符集 $t(X)$中所有事务标识符的最大公共项集。

定理 4. 如果 Z 是 X 的闭包，则有 $\sup(X) = \sup(Z)$。

证明：如果 X 是封闭的，则有 $c(X)=X$，当然就有 $\sup(X) = \sup(Z)$；如果 X 不是封闭的，则根据理解 2 关于寻找 $c(X)$的过程知道，购买 X 和购买 $c(X)$的是同一群顾客，即它们对应的事务标识符集合 T 是一样的，而 T 中的顾客数正是 X 和 $c(X)$的支持度，所以有 $\sup(X) = \sup(Z)$。

定义 6 [10]. 所有闭频繁项集的集合定义如下：

$$C = \{X \mid X \in F \text{ 且不存在 } Y \supset X, \text{使得} \sup(X) = \sup(Y)\}.$$

换句话说，X 是封闭的，若 X 的所有超集的支撑度更小，即对于所有的 $Y \supset X$，都有 $\sup(X) > \sup(Y)$。

定理 5. 若存在一个闭频繁项集 $Z \in C$ 使得项集 X 满足 $X \subseteq Z$，则 X 是频繁的。更进一步，X 的支持度为 $\sup(X) = \max\{\sup(Z) \mid Z \in C, X \subseteq Z\}$ [10]。

证明：

首先，因为 $X\subseteq Z$，所以有 $\sup(X)\ge\sup(Z)$。

其次，设 X 的闭包为某个 Z，即，根据定理 4，有 $\sup(X)=\sup(Z)$。

综上，有 $\sup(X)=\max\{\sup(Z)\mid Z\in C,X\subseteq Z\}$。

上面的定理 5 在文献[10]中有相应描述，但并没有给出标准的证明过程。

理解 3：根据上面定理，如果 X 是闭频繁项集，则其超集支持度严格比它的小；如果 X 不是闭频繁项集，则存在其超集 Z，使得 $\sup(Z)=\sup(X)$，且这样的超集 Z 就是 X 闭包。

定理 6. 对于频繁项集 X，设 $S=\{Z\mid Z\supseteq X\text{ 且 }\sup(Z)=\sup(X)\}$，$S$ 里最大(项数最多)那个超集 Z_m，就是 X 的闭包。

证明：根据 $S=\{Z\mid Z\supseteq X\text{ 且 }\sup(Z)=\sup(X)\}$，因为 $\sup(Z)=\sup(X)$，说明购买超集商品集 Z 的顾客群和购买商品集 X 的顾客群是一样的，即 $t(Z)=t(X)$。而根据理解 1，X 的闭包就顾客集合 $t(X)$ 或者 $t(Z)$ 购买的最大商品集合 $Z_m\in S$，而这个最大商品集合 Z_m 也是该顾客群购买的，即 $t(Z_m)=t(Z)=t(X)$，所以 Z_m 是 X 的闭包。

理解 4：根据定理 6，可以把支持度相等的集合归为一个等价类 $S=\{Z\mid Z\supseteq X\text{ 且 }\sup(Z)=\sup(X)\}$，其中，使得 $|Z_m|=\max(\{|Z|,Z\in S\})$ 的 Z_m 是该等价类中所有项集的闭包。

理解 5：根据定理 6 和定理 3(1)，同一个等价类里，类代表“闭包”这个项集包含了该类所有的其他项集。

定理 7：频繁项集 X 的闭包是所有包含 X 的闭频繁项集中最小(项数最少)的。

证明：

设 $Z'=\{Z\mid Z\supseteq X,\text{且不存在 }Y\supset Z\text{ 使得 }\sup(Y)=\sup(Z)\}$，那么 X 的闭包 $c(X)$ 就在 Z' 里面，假设 Z 中有闭项集 Z_1，$Z_1\supseteq X$，使得 $|c(X)|>|Z_1|$，出现下面两种情况：

(1) 如果 $c(X)\supseteq Z_1$，

则 $\sup(c(X))\le\sup(Z_1)$，又因为 Z_1 是闭项集，所以不存在 $c(X)\supseteq Z_1$ 使得 $\sup(c(X))=\sup(Z_1)$，所以有 $\sup(c(X))<\sup(Z_1)$，说明 $c(X)$ 和 Z_1 不在一个等价类，则 Z_1 不能是等价类的闭包，Z_1 不可能是 X 的闭包。

(2) 如果 $c(X)\not\supseteq Z_1$，根据理解 5，说明 Z_1 和 $c(X)$ 不在同一个等价类，则 Z_1 不能是等价类的闭包，Z_1 不可能是 X 的闭包。这里还不对。

综上，不存在闭项集 Z_1，$Z_1\supseteq X$，使得 $|c(X)|>|Z_1|$；说明，凡是有这样的闭项集 Z_1，$Z_1\supseteq X$，就有 $|c(X)|\le|Z_1|$，证毕。

注意：上面提到的项集的最大和最小，是指其包含项数的最多或者最少。

综合以上定义、定理及理解，得到如下结论：

一个二元数据库 D 上所有闭频繁项集的集合 C 是该数据库的一种紧凑表示，其中的每个闭频繁项集 X 是所有和它支持度相等、以它为闭包的商品集合的代表。

根据例 1 的商品购买数据库 D，可以得到图 1 表示的情形，其中，每个矩形小方块里面是一个频繁项集和其对应的事务标识符集合，比如商品 A 被顾客集合{1,2,4,5}里每个顾客都购买了；每个带阴影的小方块里的项集是闭频繁项集；以某个闭频繁项集为闭包的商品项集被划分到同一个不规则圈里，可以看成一类，该类的项集具有相同的支持度。比如这 4 个商品集合{A、AB、AE、ABE}所在的圈内，ABE

是闭频繁项集，其它三个商品集合都以它为闭包，ABE 就是该类的代表，这 4 个频繁项集具有相同的支持度，即 $\sup(\mathrm{A}) = \sup(\mathrm{AB}) = \sup(\mathrm{AE}) = \sup(\mathrm{ABE})$。该图是文献[10]的第 9 章的图 9-2 的一部分，我们主要关注图里的一般频繁项集、闭频繁项集以及它们相应的等价类划分，比文献[10]里关注和表达的主题少，这里主要是用来验证我们关于闭频繁项集的结论。

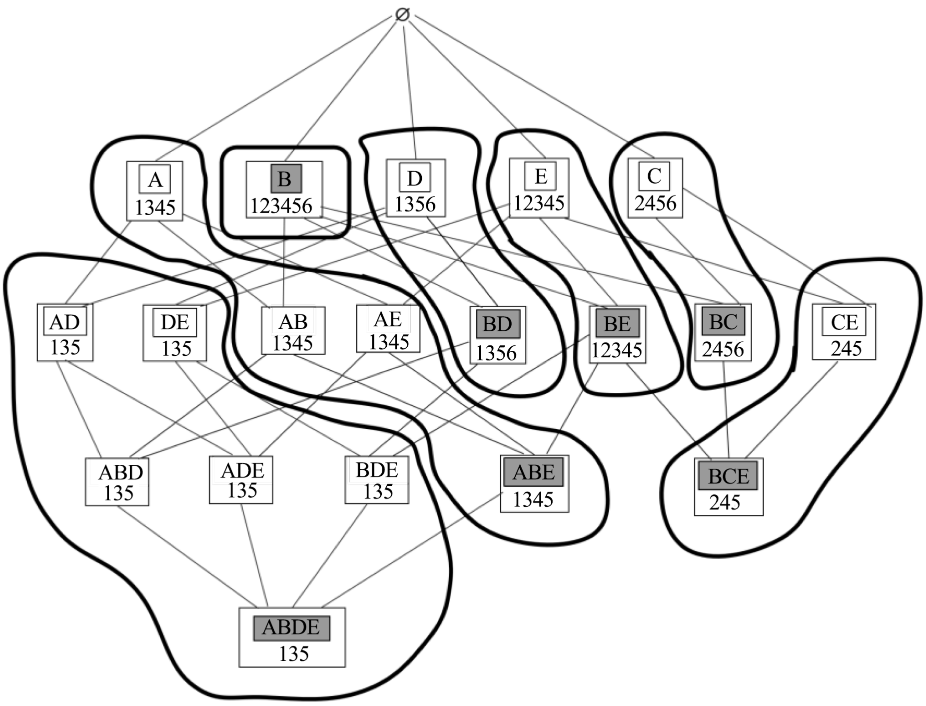

Figure 1. Schematic diagram of the closed itemset and equivalence classes for the goods purchase in Example 1
图 1. 例 1 的商品购买情况的闭项集和等价类分别示意图

3. 结论

挖掘闭频繁项集，是关联规则分析中挖掘频繁项集的一种紧凑和简洁的挖掘方式，并因此挖掘到感兴趣的关联规则。本文对闭频繁项集的定义和细节进行了梳理，其中对于闭包定义中的函数 t 和 i 进行了单调性证明；对闭包算子定义满足的性质进行了证明；对频繁项集和其闭包、以及它们的超集的支持度之间的关系进行了分析和证明，反过来根据支持度和项集之间的包含情况进行了项集的闭包的确定并给出了相应结论，从而进行了等项集等价类的划分。整个论文在购物篮关联规则分析的情景中对闭频繁项集的理论进行了透彻的分析和探究，相对于目前众多研究者们只注重挖掘闭频繁项集的算法来说，我们反过来，对其本身的理论进行探讨、从底层进行分析，希望在闭频繁项集的理论方面更进一步探索到有意义的细节，然后再促进相应挖掘算法的进展，这也是我们未来一个阶段研究思路。

基金项目

全国统计科学研究重大项目(2021LD01)。

参考文献

[1] Agrawal, R. (1994) A Fast Algorithm for Mining Association Rules in Image. 2014 *IEEE* 5*th International Conference*

on Software Engineering and Service Science, Beijing, 27-29 June 2014, 513-516. https://doi.org/10.1109/ICSESS.2014.6933618

[2] Wong, M.-H., Sze-To, H.-Y., Lo, L.Y., *et al*. (2015) Discovering Binding Cores in Protein-DNA Binding Using Association Rule Mining with Statistical Measures. *IEEE/ACM Transactions on Computational Biology & Bioinformatics*, **12**, 142-154. https://doi.org/10.1109/TCBB.2014.2343952

[3] Xiang, Y., Shuai, C., Li, Y. and Zhang, Y. (2022) Information Reconstruction of Student Management Work Based on Association Rules Mining. *Computational Intelligence and Neuroscience*, **2022**, Article ID: 2318515. https://doi.org/10.1155/2022/2318515

[4] Shrestha, A., Zikos, D. and Fegaras, L. (2021) An Annotated Association Mining Approach for Extracting and Visualizing Interesting Clinical Events. *International Journal of Medical Informatics*, **148**, 104366. https://doi.org/10.1016/j.ijmedinf.2020.104366

[5] Hu, K., Qiu, L., Zhang, S., *et al*. (2023) An Animal Dynamic Migration Optimization Method for Directional Association Rule Mining. *Expert Systems with Applications*, **211**, 118617. https://doi.org/10.1016/j.eswa.2022.118617

[6] Bezdek, J.C. (1974) Numerical Taxonomy with Fuzzy Sets. *Journal of Mathematical Biology*, **7**, 57-71.

[7] Bezdek, J.C. (1974) Cluster Validity with Fuzzy Sets. *Journal of Cybernetics*, **3**, 58-72. https://doi.org/10.1080/01969727308546047.

[8] 李广璞，黄妙华. 频繁项集挖掘的研究进展及主流方法[J]. 计算机科学, 2018, 45(11): 1-11, 26. https://www.jsjkx.com/CN/Y2018/V45/I11A/1

[9] 谭军，卜英勇，杨勃. 一种高效的闭频繁模式挖掘算法[J]. 计算机工程与应用, 2010, 46(6): 130-132. https://doi.org/10.3778/j.issn.1002-8331.2010.06.037

[10] [美]穆罕默德·扎基，小瓦格纳·梅拉. 数据挖掘与分析-概念与算法[M]. 吴诚堃，译. 北京：人民邮电出版社, 2017.

[11] 毛伊敏，陈志刚. 在线挖掘数据流闭频繁项集的高效算法[J]. 计算机科学, 2013, 40(2): 229-234. https://doi.org/10.3969/j.issn.1002-137X.2013.02.050

[12] 柴玉梅，张卓，王黎明. 基于频繁概念直乘分布的全局闭频繁项集挖掘算法[J]. 计算机学报，2012，35(5): 990-1001. https://doi.org/10.3724/SP.J.1016.2012.00990

[13] Aryabarzan, N. and Minaei-Bidgoli, B. (2021) NEclatClosed: A Vertical Algorithm for Mining Frequent Closed Itemsets. *Expert Systems with Applications*, **174**, 114738. https://doi.org/10.1016/j.eswa.2021.114738

[14] Li, B., Pei, Z., Qin, K., *et al*. (2019) TT-Miner: Topology-Transaction Miner for Mining Closed Itemset. *Institute of Electrical and Electronics Engineers* (*IEEE*), **7**, 10798-10810. https://doi.org/10.1109/access.2018.2888627

[15] Wu, Y., Wang, M., Chu, W., *et al*. (2023) Organization Preference Knowledge Acquisition of Multi-Platform Aircraft Mission System Utilizing Frequent Closed Itemset Mining. *Aerospace*, **10**, Article 166. https://doi.org/10.3390/aerospace10020166

[16] Liu, J.Q., Ye, Z.S., *et al*. (2021) Efficient Strategies for Incremental Mining of Frequent Closed Itemsets over Data Streams. *Expert Systems with Applications*, **191**, 116220. https://doi.org/10.1016/j.eswa.2021.116220

[17] Pasquier, N., Bastide, Y., Taouil, R. and Lakhal, L. (1999) Efficient Mining of Association Rules Using Closed Itemset Lattices. *Information Systems*, **24**, 25-46. https://doi.org/10.1016/S0306-4379(99)00003-4.

[18] Ganter, B. and Wille, R. (1999) Formal Concept Analysis: Mathematical Foundations. Springer-Verlag, Berlin. https://doi.org/10.1007/978-3-642-59830-2