

文章编号: 2095-2163(2019)04-0051-04

中图分类号: TP311.13

文献标志码: A

# 基于关联规则的医疗器械预配货模板的设计与实现

贺紫璐

(东华大学 计算机科学与技术学院, 上海 201620)

**摘要:** 随着医疗器械供应链管理朝着信息化和透明化的趋势发展,市面上出现了越来越多医疗器械供应链平台。但这些供应链平台的仓储模式比较繁琐,需要人工确认订单所需商品信息,大量订单产生后使仓库配货造成人工失误的几率越来越大。由于每笔订单中相同手术部位所需器械大部分相同,本系统分析了业务订单中客户长期搭配购买的医疗器械,通过使用 Apriori 算法和泊松分布计算商品之间和商品数量之间的关联关系,从而设计了多套预配货模板,这些模板使仓库人员可以提前准备与模板相匹配的预配货箱子,从而提高仓库的配货效率。

**关键词:** 医疗器械供应链; 预配货; 关联规则; Apriori 算法

## Design and implementation of medical device pre-distribution templates based on association rules

HE Zijun

(School of Computer Science and Technology, Donghua University, Shanghai 201620, China)

**[Abstract]** With the development of medical device supply chain management towards informatization and transparency, more and more medical device supply chain platforms appear on the market. However, the warehousing mode of these supply chain platforms is rather complicated, which requires manual confirmation of commodity information required for orders. After a large number of orders are generated, the probability of manual errors in warehouse allocation becomes greater and greater. Because most of the instruments required for the same surgical site in each order are the same, this system analyses the medical devices that customers purchase in long-term collocation in business orders. By using Apriori algorithm and Poisson distribution, the relationship between commodities and the quantity of commodities is calculated, so a number of pre-dispatch templates are designed. These templates enable warehouse personnel to prepare the pre-dispatch distribution boxes matching with the template in advance, thereby improving the efficiency of warehouse distribution.

**[Key words]** medical device supply chain; pre-distribution; association rules; Apriori algorithms

## 0 引言

中国医疗器械产业整体发展势头迅猛,为医疗器械专业物流带来了广阔的发展机遇。但是,由于器械设备的专业性、产品以及流通渠道的特殊性,医疗器械物流一直未能受到广泛重视,专业的第三方医疗器械物流企业数量寥寥无几,主要原因在于行业监管松散和技术应用落后<sup>[1]</sup>。随着信息技术及互联网的飞速发展,各种信息数据正在快速的增长,电子商务也在经历飞速的发展,医疗器械电商配送渐成体系<sup>[2]</sup>,订单数量急剧增长。仓库中需要大量重复利用的订单使配货流程变得更加繁琐,急需一套能够帮助企业节省配货效率的方法。

本文利用数据挖掘相关技术设计一套企业预配货模板,利用关联规则分析器械之间的关联关系,而 Apriori 算法是数据挖掘中进行关联规则挖掘的经

典算法。如文献[3]利用 Apriori 算法挖掘高校学生成绩,文献[4]利用该算法挖掘高校 Web 日志,文献[5]分析航空设备故障关联关系,文献[6]利用数据挖掘技术分析研究了电力系统二次设备缺陷等等。

利用关联规则挖掘技术可以发现许多隐藏在数据背后的信息,如发现订单的医疗器械之间相互关联关系。本文利用关联挖掘技术计算出订单中相关的预配货模板,使得系统管理员可以在供应链平台中添加预配货模板,从而使仓库人员提前通过预配货模板在仓库中装配好与预配货模板对应的多个预配货箱子,使得每次发货时直接将预配货箱子运送给客户,达到减少配货时间、提高发货效率的功能。

## 1 相关分析

### 1.1 关联规则

如果 2 个或多个变量的取值之间存在某种规律

作者简介: 贺紫璐(1994-),女,硕士研究生,主要研究方向:大数据、关联规则。

收稿日期: 2019-04-03

性,则称为关联。关联规则是寻找在同一个事件中  
出现不同项的相关性,在机器学习研究领域,学者们  
已经提出多种关联规则的挖掘算法,如 Apriori、FP-  
growth、Eclat 算法等,本文主要使用的是 Apriori 算法。

### 1.1.1 关联规则相关定义

**定义 1** 项、项集和  $k$ -项集,事务和事务集。假设  $I = \{i_1, i_2, \dots, i_m\}$  表示所有项的集合,  $I$  是  $m$  个不同项的集合,每个  $i_k (k = 1, 2, \dots, m)$  称为项,  $I$  为项的集合,即项集;包含  $k$  个项的项集称为  $k$ -项集。假设  $D = \{T_1, T_2, \dots, T_n\}$  表示所有事务的集合,每个事务  $T_i$  由项构成,即  $T_i \in I$ 。假设以 0 和 1 表示项是否在事务中出现,具体表示见表 1<sup>[7]</sup>:

表 1 事务数据库  $D$  中的二元 0/1 表示

Tab. 1 Binary 0/1 representation in transaction database  $D$

TID	$i_1$	$i_2$	.....	$i_m$
$T_1$	1	0	.....	1
$T_2$	1	1	.....	0
.....	.....	.....	.....	.....
$T_n$	0	1	.....	0

**定义 2** 关联规则。关联规则可表示成  $X \Rightarrow Y$  的蕴含式,其中  $X \in I, Y \in I$ ,且  $X \cap Y = \emptyset$ 。

**定义 3** 关联规则的支持度。关联规则的支持度是指事务数据库  $D$  中,蕴含有  $X$  以及  $Y$  事务占据全部事务的百分比,即:

$$\text{support}(X \Rightarrow Y) = \frac{D \text{ 中包含项集 } X \cup Y \text{ 的事务数}}{D \text{ 中事务总数}} \\ = P(X \cup Y)。$$

一般将其看成是最小支持度,记为  $\text{min\_sup}$ 。

**定义 4** 置信度。关联规则  $X \Rightarrow Y$  置信度是指  $D$  中包含  $X$  的事务中同时也包含  $Y$  事务的百分比,即为:

$$\text{confidence}(X \Rightarrow Y) = \frac{\text{support}(X \Rightarrow Y)}{\text{support}(X)} \\ = \frac{P(X \cup Y)}{P(X)} = P(Y | X)。$$

一般将其看成是最小置信度,记为  $\text{min\_conf}$ 。

**定义 5** 频繁项集。满足最小支持度的项集即为频繁项集,记为  $L_k$ 。

**定义 6** 强关联规则。强关联规则指的是满足最小支持度和最小置信度的规则。关联规则挖掘过程分为 2 个步骤:找出事务数据库中所有频繁项集;由频繁项集产生强关联规则。

### 1.1.2 Apriori 算法

Apriori 算法<sup>[3]</sup>是关联规则中经典算法之一。要

想获得频繁  $k$  项集,必须通过频繁  $k - 1$  项集迭代得到,以此类推,得到全部项集。Apriori 算法主要步骤如下:

#### (1) 连接

要得到频繁项集  $L_k$ , 首先需要计算出候选项集  $C_k = L_{k-1} \times L_{k-1}$ 。假设  $l_1$  和  $l_2$  是  $L_{k-1}$  内的 2 个项集,且  $l_1$  和  $l_2$  中有  $(k - 2)$  个项相同。设相同项为  $\{l[1], l[2], \dots, l[k - 2]\}$ , 不同项分别为  $l_1[k - 1]$  和  $l_2[k - 1]$ , 则表示  $l_1$  和  $l_2$  可连接,连接结果为  $C_k = \{l[1], l[2], \dots, l[k - 2], l_1[k - 1], l_2[k - 1]\}$ ;

#### (2) 剪枝

由于  $C_k$  是  $L_k$  的超集,故  $C_k$  中可能包含由  $L_k$  自连接产生的非频繁项集,需将这些非频繁项集删除,才可得到  $L_k$ 。主要是利用 Apriori 算法的先验性质:若有某一候选项集  $t \in C_k$ ,且  $t$  的某个子集  $s \notin L_k$ ,即  $s$  不是频繁项集,故将  $t$  从  $C_k$  中删除;

#### (3) 产生强关联规则

通过(1)和(2)步骤找出事务数据库  $D$  中的频繁项集  $L_k$ , 通过以下步骤得到满足最小支持度和最小置信度的关联规则:

①计算出每个频繁项集  $l$  的所有非空子集;

②对于每个  $l$  的非空子集  $p$ , 若  $\frac{\text{support}(l)}{\text{support}(p)} \geq$

$\text{min\_conf}$ , 则称  $p \Rightarrow (l - p)$  为强关联规则。

## 1.2 泊松分布

设随机变量  $X$  所有可能取值为  $0, 1, 2, \dots$ , 而取各个值的概率是:

$$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \dots$$

其中,  $\lambda > 0$  是常数。则称  $X$  是服从参数  $\lambda$  的泊松分布<sup>[8]</sup>, 记为  $X \sim \pi(\lambda)$ 。

## 2 基于关联规则的医疗器械预配货模板的设计

本功能模块通过分析大量相同手术部位的订单信息,发现大部分订单会搭配使用一些医疗器械。经过分析与整理每一笔订单里经常会搭配出现的医疗器械,将这些医疗器械组合起来,形成一个预配货模板。本文主要通过 3 个步骤设计企业预配货模板:

(1)给每种器械进行编号,通过关联规则计算出搭配使用的医疗器械的编号;

(2)通过泊松分布计算出相应器械数量存在于订单中的大概率区间;

(3)通过关联规则分析各器械数量的大概率区

间是否相关,若相关则各器械取能够满足大部分订单的数量;若无关则取其中一个器械的大概率区间,分析出与之相关器械的取值区间,最后取这些区间里的中位数为预配货模板里对应器械的数量。

**2.1 利用 Apriori 算法找出搭配使用的器械编号**

本文中的项指的是  $n$  种器械的编号,即  $1 \sim n$ ,故项集  $I = \{1, 2, \dots, n\}$ ; 每笔交易称为一个事务  $T$ 。如若第  $i$  笔交易中使用了 1 号和 3 号器械,则事务数据库  $D_1$  中  $T_i = \{1, 3\}$ , 设置最小支持度阈值 ( $\text{min\_sup}$ ) 为 0.8。事务数据库  $D_1$  相应的二元 0/1 表示见表 2。

表 2 事务数据库  $D_1$  中的二元 0/1 表示

Tab. 2 Binary 0/1 representation in transaction database  $D_1$

TID	1	2	3	.....	$n$
$T_1$	1	0	0	.....	1
$T_2$	1	1	1	.....	0
.....	.....	.....	.....	.....	.....
$T_i$	1	0	1	.....	0
.....	.....	.....	.....	.....	.....
$T_n$	0	1	0	.....	0

由第 1.1.1 小节可知,通过计算使得

$$\text{support}(A) = \frac{D \text{ 中包含项集 } A \text{ 的事务数}}{D \text{ 中事务总数}} \geq \text{min\_sup},$$

即  $A$  为频繁项集,使用 Apriori 算法逐层的迭代计算项集  $I$  在事务数据库  $D_1$  中的频繁 1 项集,频繁 2 项集, ..., 频繁  $k$  项集,每次迭代计算结束后通过先验性质剔除部分项集,得到最终频繁  $k$  项集  $L_k$ 。(  $L_k = \{l_1, l_2, \dots, l_k\}$  里的每一项  $l_i (i \in [1, k], \text{且 } i \text{ 为整数})$ , 表示一个预配货模板中所包含器械的编号。

**2.2 利用泊松分布和 Apriori 算法得出每个模板中所有器械的数量**

假设通过本文第 2.1 小节,已找出所有订单中搭配使用的器械组合,以某个预配货组合  $l_i (i \in [1, k], \text{且 } i \text{ 为整数}) = \{x, y, z\}$  为例。

**2.2.1 利用泊松分布计算出模板中器械数量的概率分布区间**

分别计算出当  $x, y$  和  $z$  这 3 种类型器械进行组合时,每种器械数量的概率分布:

(1)统计包含该种组合的所有订单中  $x$  的数量,设为数组  $C = [x_1, x_2, \dots, x_n]$ ,取  $C$  中所有数的平

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n};$$

(2)将  $\bar{X} \approx \lambda$ ,则关于  $x$  编号的器械的泊松分布为:

$$P_x\{X = k\} = \frac{\bar{X}^k e^{-\bar{X}}}{k!}, k = x_1, x_2, \dots, x_n;$$

(3)求出当  $P_x\{X = k\} \geq 0.8$  时的所有  $x$  的取值,该范围为器械  $x$  在所有订单中的取值范围,记为  $[x_i, x_j]$ ;

(4)同理计算可得器械  $y$  和  $z$  在所有订单中的出现概率大于等于 80% 时的取值范围,记为  $[y_i, y_j]$  和  $[z_i, z_j]$ 。

**2.2.2 利用 Apriori 算法计算其余器械的数量取值**

(1)将所有同时出现  $x, y, z$  的订单设成新的事务数据库  $D_2$ , 器械  $x, y$  和  $z$  的个数为每个事务  $T_i = \{a, b, c\}$ ,  $i$  为某个订单,  $a, b$  和  $c$  分别对应器械  $x, y$  和  $z$  在每个订单的数量;

(2)设置最小支持度阈值 ( $\text{min\_sup}$ ) 和最小置信度阈值 ( $\text{min\_conf}$ ) 均为 80%。若事务  $T_i$  能够满足  $a \in [x_i, x_j], b \in [y_i, y_j], c \in [z_i, z_j]$  则  $T_i = \{a, b, c\}$ , 事务数据库  $D_2$  相应的二元 0/1 表示见表 3;

表 3 事务数据库  $D_2$  中的二元 0/1 表示

Tab. 3 Binary 0/1 representation in transaction database  $D_2$

TID	$a$	$b$	$c$
$T_1$	1	0	0
$T_2$	1	1	1
.....	.....	.....	.....
$T_i$	1	1	1
.....	.....	.....	.....
$T_n$	0	1	0

(3)设新的项集  $I_1 = \{a, b, c\}$ , 通过 Apriori 算法确认该项集  $I_1$  是否能够满足  $\text{support}(a \Rightarrow b) \geq 0.8$  以及是否满足  $\text{support}(a \Rightarrow c) \geq 0.8$ , 若满足则通过上述泊松分布求出  $P_x\{X = k\} \geq 0.9$  时,将每次计算得出  $x$  的最小值分别赋给  $a, b, c$  成为模板中器械  $x, y, z$  分别对应的数量。若不满足进行第(4)步;

(4)通过 Apriori 算法求出满足  $\text{support}(a \Rightarrow b) \geq 0.8$  和  $\text{support}(a \Rightarrow c) \geq 0.8$  的  $b$  和  $c$  的区间。将该区间的中位数作为器械  $y$  和  $z$  对应的数量。

**3 实现**

**3.1 基于关联规则的医疗器械预配货模板的实现**

本文使用 python 实现 Apriori 算法计算预配货模板,主要代码如图 1 所示。

图 1 中有一个关键函数 aprioriGen,主要用于生成候选项集,该函数主要功能为:

```

def apriori(datasets, minSupport = 0.8):
    C1 = createC1(datasets)
    D = list(map(set, datasets))
    L1, supportData = scan0(D, C1, minSupport) # 单项最小支持度判断 0.8, 生成L1
    L = [L1]
    k = 2
    while (len(L[k-1]) > 0): # 创建包含更大项集的更大列表, 直到下一个大的项集为空
        Ck = aprioriGen(L[k-1], k) # 生成Ck
        Lk, supk = scan0(D, Ck, minSupport) # 得到Lk
        supportData.update(supk)
        L.append(Lk)
        k += 1
    return L, supportData

```

图1 Python实现Apriori算法主要代码图

Fig. 1 Main code diagram of Apriori algorithms implemented by Python

def aprioriGen(Lk, k): # 组合, 向上合并

# creates Ck 参数: 频繁项集列表 Lk 与项集元素个数 k

retList = []

lenLk = len(Lk)

for i in range(lenLk):

for j in range(i + 1, lenLk): # 两两组合遍历

L1 = list(Lk[i]][:k-2]; L2 = list(Lk[j]][:k-2]

L1.sort(); L2.sort();

if L1 == L2: # 若 2 个集合的前 k-2

个项相同时, 则将 2 个集合合并

retList.append(Lk[i] | Lk[j]) #

set union

return retList

### 3.2 添加预配货模板的实现

得到预配货模板信息后, 系统管理员将相应的信息录入系统, 本部分功能主要是对预配货模板的增删改查, 主要涉及的类见表 4。

表 4 预配货模板功能模块主要类

Tab. 4 Main classes of function modules for pre-distribution templates

类名	文件名
控制层类	PrematchBoxInfoController.java
	PrematchTemplateInfoController.java
	PrematchBoxInfoService.java
业务逻辑类	PrematchTemplateInfoService.java
	PrematchBoxInfoServiceImpl.java
	PrematchTemplateInfoServiceImpl.java
持久对象类	PrematchBoxInfo.java
	PrematchTemplateInfo.java
数据持久类	PrematchBoxInfoRepository.java
	PrematchTemplateInfoRepository.java

首先预配货模板汇总页面显示所有的预配货模板, 该功能主要是通过 PrematchTemplateInfoController 类的 findAllByPage 方法实现, 使用 PC 端创建新模板界面如图 2 所示。该页面保存模板的功能主要通过 PrematchTemplateInfoController 类的 saveTemplate 方法实现, 查看模板信息是通过 PrematchTemplateInfoController 类的 finddetail 方法实现。

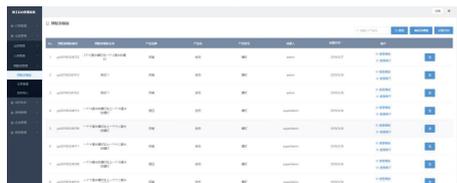


图 2 预配货模板汇总页面图

Fig. 2 Pre-distribution template summary page interface diagram

## 4 结束语

本文通过分析企业大量订单信息, 为减少企业配货时间, 提高企业配货效率, 结合数据挖掘中关联挖掘的相关知识, 利用 Apriori 算法和泊松分布为企业计算出与订单相匹配的预配货模板, 使企业在下单之前通过该模板装配好对应器械和数量的预配货箱子, 从而通过订单中商品信息直接进行发货操作。本方法经过用户的实际使用能够满足业务需求, 并为其它类似系统的研发提供了借鉴。

## 参考文献

- [1] 崔忠付. 我国医疗器械物流发展问题与趋势分析[J]. 物流技术与应用, 2015, 20(11): 50-51.
- [2] 叶楠. 医疗器械电商转型的全渠道战略—康复之家案例研究[D]. 北京: 北京外国语大学, 2017.
- [3] 吴小东, 曾玉珠. 基于 Apriori 算法的高校学生成绩数据挖掘[J]. 廊坊师范学院学报(自然科学版), 2019, 19(1): 31-36.
- [4] 王春玲, 李川, 李想. 基于 Apriori 算法的高校 Web 日志挖掘系统构建[J]. 中国林业教育, 2019, 37(2): 22-26.
- [5] 陈秀秀, 刘凯, 马双涛, 等. 基于 Apriori 算法的航空设备故障关联分析[J]. 山东师范大学学报(自然科学版), 2019, 34(1): 48-53.
- [6] 皇甫汉聪, 肖招娣. 基于 Apriori 算法的电力系统二次设备缺陷数据挖掘与分析研究[J]. 电子设计工程, 2019, 27(5): 6-9, 15.
- [7] TAN Pangning, STEINBACH M, KUMAR V. 数据挖掘导论[M]. 范明, 范宏建, 译. 北京: 人民邮电出版社, 2010.
- [8] 赵瑛. 关于泊松分布及其应用[J]. 辽宁省交通高等专科学校学报, 2009, 11(2): 77-78.