Service Science and Management 服务科学和管理, 2019, 8(6), 234-239
Published Online November 2019 in Hans. http://www.hanspub.org/journal/ssem
https://doi.org/10.12677/ssem.2019.86035

Hans汉斯

A Study on Dish-Items Configuration Rules Based on the Apriori Arithmetic

Taotao Yu, Gefu Zhang*, Zhaohui Hu

Economic Management and Law School, University of South China, Hengyang Hunan
Email: *gzjkfu@qq.com

Received: Sep. 4th, 2019; accepted: Oct. 25th, 2019; published: Nov. 1st, 2019

Abstract

Shops which win with sales volume, must avoid casualness and blindness to control stocks. With a whole year clean dishes sale data of chain shops, based on the Apriori association rule arithmetic, applying IBM modeler platform, it has constructed an association rule mining model of clean dishes in chain shops. 24 association rules, highly close to the dietary pattern of residents, have been gotten with lift bigger than 3, which prove the scientific of this arithmetic model. These rules, drawn from tens of thousands of transaction records, can optimize dish items configuration and guide to stack them on shelves.

Keywords

Apriori, Recommendation Arithmetic, Dish-Items Configuration

基于Apriori算法的菜品配置规则研究

余滔滔，张革伕*，胡朝晖

南华大学经济管理与法学学院，湖南 衡阳
Email: *gzjkfu@qq.com

收稿日期：2019年9月4日；录用日期：2019年10月25日；发布日期：2019年11月1日

摘　要

门店以销量取胜，需避免随意性与盲目性来控制库存。本文研究了净菜连锁门店一年的销售数据，基于Apriori关联规则算法，应用IBM的Modeler平台，构造出连锁门店的净菜关联规则挖掘模型。挖掘到了24条提升度大于3的关联规则，与居民的烹饪规律高度吻合，证明了算法模型的科学性。规则从数万条

*通讯作者。

文章引用: 余滔滔, 张革伕, 胡朝晖. 基于 Apriori 算法的菜品配置规则研究[J]. 服务科学和管理, 2019, 8(6): 234-239.
DOI: 10.12677/ssem.2019.86035

交易记录中提取，能够优化门店的菜品配置，也能指导菜品上架。

关键词

Apriori，推荐算法，菜品配置

1. 引言

随着消费的转型升级，百姓饮食结构在发生改变，在菜品的选购上趋个性化，同时量小而精、美味、健康成为消费准则。菜店等供货商如何有效组织菜品上架以满足消费者需求，成为各店铺在激烈竞争中取胜的法宝。薄治禹[1]以快餐业为代表，应用卡方方法来研究饮食行业的关联规则问题，提出服务员在就餐者点菜时，重点将“肉汉堡”与“可乐”、“肉丝汤”与“蛋炒饭”这 2 组餐饮产品向就餐者推荐。从高校食堂的就餐系统中提取数据来研究，王楚瑜等[2]应用 Apriori 算法，研究了“素菜－素菜、素菜－荤菜之间的销量关联性”，发现“猪排盖饭与优酸乳”具强关联特征。李筠等[3]从药效角度，研究了吃药期间饮食搭配下的禁忌规则。在大数据背景下，越来越多的研究人员开始研究消费推荐系统，这种关联既包括人与产品之间的关联，也包括产品之间的关联。邱京伟等[4]研究了粒关联规则挖掘算法，探索顾客与菜品间的粒关联规则，来提高菜品推荐命中率。张奥多等[5]应用 FP-tree 算法，综合商家利益最大化、热销菜品等因素，来构建推荐综合评分模型，从而实现推荐。利用用户的手机 APP 来获取 LBS 地理位置信息，向其推荐附近的餐饮企业及其招牌菜，有比较成功的应用[6]。

洗净菜门店向周边居民提供净菜，然而采购太多容易浪费，菜品搭配不当易致过剩、增加库存成本，使消费者体验差。显然，研究菜品配置和上架规则非常重要。

2. Apriori 算法

关联规则学术界也称之为“购物篮规则”，用以找出购物篮内货物品项之间的可能关联性，为商家在采购货品、上架摆放提供决策方案，有助于精准流通、提高销量。经典的场景如：在尿不湿的现场摆放啤酒。尿不湿时婴儿用品，啤酒通常是大男人的专属，这两个对象的关联特征在购物篮内被挖掘出来了。找出这种前项与后项关联规则的方法，有 Apriori、灰色关联法和 FP-tree 等等，其中最经典的是 Apriori 算法。

Apriori 算法的基本思想是：从购买事务数据中提取商品项，对商品项进行排序，构建商品项二维矩阵，然后：(1) 进行非重复的两两前后连接，(2) 对连接所得项进行裁剪，所剩下的满足一定置信度、支持度、提升度要求的连接项，即称之为规则项。例如：有商品项 A、B、C、D，都有可能出现在购物篮中。连接项将产生 AB、AC、AD、BC、BD、CD 和 ABC、ABD、$ABCD$、BCD，从事务发生的角度，这些连接项说明了同时发生的频次。Apriori 算法就用支持度来表示项目发生(出现)的概率，用置信度来描述项目发生的相对概率。例如，已知项集的支持度计数为频次模式，则规则 $A \Rightarrow B$ 的支持度和置信度很容易从所有事务计数、项集 A 和项集 $A \cup B$ 的支持度计数推出，计算方法如下式(1)和(2)。

$$Support\left(A \Rightarrow B\right) = \frac{A,B\text{同时发生的事务个数}}{\text{所有事务个数}} = \frac{Support\_count\left(A \cup B\right)}{Total\_count\left(A\right)} \tag{1}$$

$$Confidence(A \Rightarrow B) = P(A \mid B) = \frac{Support(A \cup B)}{Support(A)} = \frac{Support\_count(A \cup B)}{Support\_count(A)} \quad (2)$$

通常会约定规则的最小支持度和最小置信度值，阈值大小由用户来确定，通常与数据量大小有关，例如10万条与1000万条记录的最小支持度同设置为3%，意义差距很大。一条规则能不能用于指导实践，还需要使用所谓的提升度(Lift)值来判断，如下式(3)来计算。

$$\text{lift}(A \Rightarrow B) = \frac{P(AB)}{P(A)P(B)} \quad (3)$$

例如在事务中，用户在无需推荐其他项购物篮中有B项的概率为50%，而规则 $A \Rightarrow B$ 提出的先购买 A 再购买 B 的概率为45%，两相比较，规则 $A \Rightarrow B$ 实则没有任何意义，推荐变成了费力不讨好。通常约定提升度需要达到3以上才认为采用规则是可取的，而IBM的Modeler同时还会用部署能力来描述使用规则的可行性。

3. 基于 Apriori 的菜品配置规则

本文研究连锁净菜门店的客户购物篮内情况，从菜品搭配形成的规则来看当地居民的饮食规律，以证明方法的科学有效，探讨规则的决策价值。

3.1. 数据准备与预处理

研究的连锁店采用了用友“Tplus”系统，数据库系统为SQL SERVER。从数据库中导出同为衡阳市区内三家门店2015~2016年间销售数据，导入到EXCEL表中，数据行数超过10万条。剔除其中销售净菜名字段无效、为空的记录，剔除交易销售额为0的记录，共获得有效交易记录65355条。每条交易记录数据包含交易单子号、一项销售的净菜名及其数量、价格，一个单子号有多条记录，包含多项净菜，为一次交易事务。本文净菜是指清洗干净、切好、包好的菜。

图1描述了在IBM Modeler中的数据处理过程。IBM公司的Modeler为专业数据挖掘平台，原名为SPSS Clementine。数据经过了合并、商品项矩阵表构造、汇总交易和矩阵表字段变换等环节。下面说明几个关键步骤：

(1) 在“合并”节点，将交易记录中的商品编码替换为菜品名称。在关系型数据库中，交易记录中只有商品编码，而规则要明了必须使用菜品名。

(2) 根据挖掘目标，在过滤与区分节点，过滤不需要的字段和交易金额为0的记录、以及单次交易额超过1000元的批发出库数据，只留下交易记录单号和交易菜品名称。每条交易单号和菜品名形成唯一事务记录。

(3) 在“重新结构化”节点，将商品名排序后，构造二维商品名矩阵。在数据库内菜品名达到1000多个，但实际销售的菜品只有170多种，因而首先需要构建已销售菜品项，减少无效连接。

(4) 在“汇总”节点，按单号进行菜品汇总，出现在每个单子中的菜品数量累加。

(5) 在“分区”节点，采用Modeler提供“训练与测试”分区方法来对所有数据按分区，按照50:50的方法来分配数据。

(6) 在“填充”节点，把一个交易单号中出现多次的商品进行0/1处理。

(7) 在“导出2”节点，进行“标志化”数据处理，将1的数据转换为“T”，否则为“F”。

3.2. 建模

Modeler平台采用可视化的数据流模式，提供了常见的数据挖掘模型，包括决策树、神经网络、关联

规则、聚类和支持向量机等模型。Modeler 的 Apriori 关联规则算法是封装的，其算法如同前述的原理。可以直接从事务型数据开始挖掘，也可从标志型数据开始进行计算，计算的过程是封装的，Modeler 提供了几处参数调整来改善模型的性能。在选择数据模型以后，Modeler 一般先采用训练数据来形成业务模型，然后再用所形成的模型来进行测试与挖掘。Modeler 提供了多种方法来调整模型的有效性和性能。在连接中，对于前项的限制，本研究采用了允许前项为 0 (允许没有前项的规则)和最大前项数为 3；仅允许标志变量为真值。在最低支持度和置信度的设置上，经多次测试，发现在大数据量情况下，即使只有个位数的支持度，仍然具有实际意义。由于实际交易数据记录数超过了 2 万 5，本研究只选用了 3%的最低支持度，意味着一年的时间内某商品项至少被卖出 750 次。图 2 描述了模型所使用的菜品项字段数 177 个，支持度、置信度阈值分别为 3%和 5%。

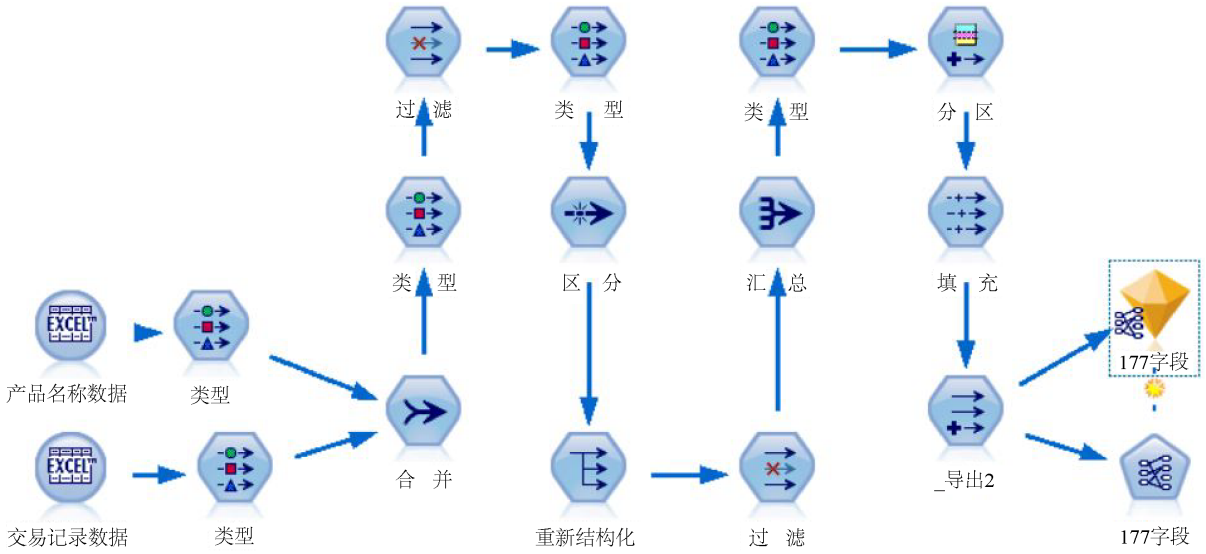
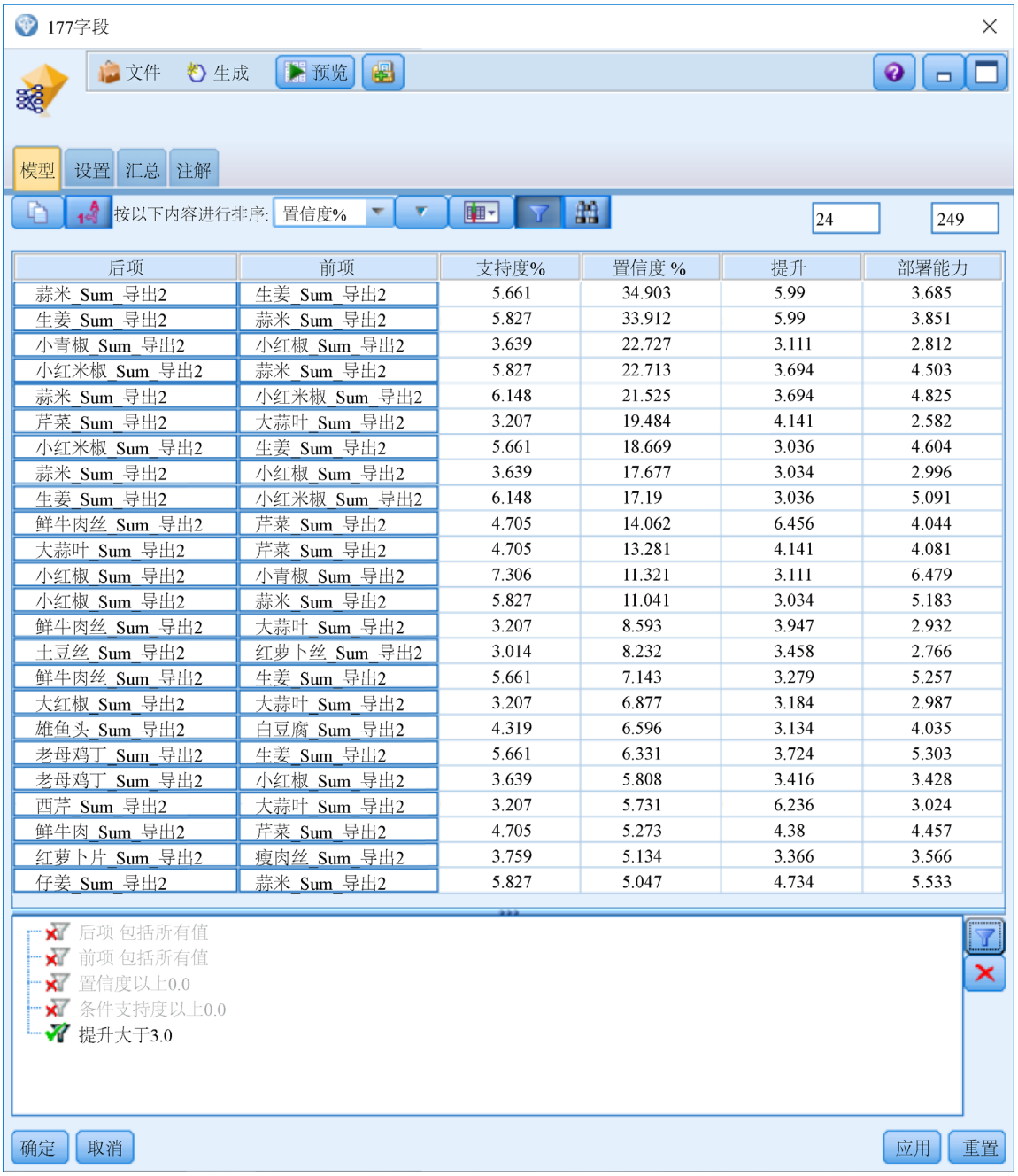


Figure 1. Data handling process
图 1. 数据处理过程

Figure 2. Apriori Model
图 2. Apriori 模型

3.3. 模型运行结论

按照上述流程，可得到 249 条规则，在默认情况下，“提升度”为 1。实践中，一般提取“提升度”大于 3 的规则。Modeler 提供了基于提升度的过滤方法，图 3 显示了“提升度最小值”为 3 时所得的规则集。Modeler 除了提供用于判断采用规则所带来的影响力即提升度，也提供了“部署能力”来判断采用规则所带来的效果空间。尽管概率值不大，但是对于数以 5 万计的交易，购买频率仍然是非常高的，例如 500 次每年，意味着每年可销售 500 件，如每件平均 4 元，单品年销售额达到 2000 元。



后项	前项	支持度%	置信度 %	提升	部署能力
蒜米_Sum_导出2	生姜_Sum_导出2	5.661	34.903	5.99	3.685
生姜_Sum_导出2	蒜米_Sum_导出2	5.827	33.912	5.99	3.851
小青椒_Sum_导出2	小红椒_Sum_导出2	3.639	22.727	3.111	2.812
小红米椒_Sum_导出2	蒜米_Sum_导出2	5.827	22.713	3.694	4.503
蒜米_Sum_导出2	小红米椒_Sum_导出2	6.148	21.525	3.694	4.825
芹菜_Sum_导出2	大蒜叶_Sum_导出2	3.207	19.484	4.141	2.582
小红米椒_Sum_导出2	生姜_Sum_导出2	5.661	18.669	3.036	4.604
蒜米_Sum_导出2	小红椒_Sum_导出2	3.639	17.677	3.034	2.996
生姜_Sum_导出2	小红米椒_Sum_导出2	6.148	17.19	3.036	5.091
鲜牛肉丝_Sum_导出2	芹菜_Sum_导出2	4.705	14.062	6.456	4.044
大蒜叶_Sum_导出2	芹菜_Sum_导出2	4.705	13.281	4.141	4.081
小红椒_Sum_导出2	小青椒_Sum_导出2	7.306	11.321	3.111	6.479
小红椒_Sum_导出2	蒜米_Sum_导出2	5.827	11.041	3.034	5.183
鲜牛肉丝_Sum_导出2	大蒜叶_Sum_导出2	3.207	8.593	3.947	2.932
土豆丝_Sum_导出2	红萝卜丝_Sum_导出2	3.014	8.232	3.458	2.766
鲜牛肉丝_Sum_导出2	生姜_Sum_导出2	5.661	7.143	3.279	5.257
大红椒_Sum_导出2	大蒜叶_Sum_导出2	3.207	6.877	3.184	2.987
雄鱼头_Sum_导出2	白豆腐_Sum_导出2	4.319	6.596	3.134	4.035
老母鸡丁_Sum_导出2	生姜_Sum_导出2	5.661	6.331	3.724	5.303
老母鸡丁_Sum_导出2	小红椒_Sum_导出2	3.639	5.808	3.416	3.428
西芹_Sum_导出2	大蒜叶_Sum_导出2	3.207	5.731	6.236	3.024
鲜牛肉_Sum_导出2	芹菜_Sum_导出2	4.705	5.273	4.38	4.457
红萝卜片_Sum_导出2	瘦肉丝_Sum_导出2	3.759	5.134	3.366	3.566
仔姜_Sum_导出2	蒜米_Sum_导出2	5.827	5.047	4.734	5.533

Figure 3. Effect of Apriori vegetable recommendation model
图 3. Apriori 菜品推荐模型效果

4. 菜品配置推荐

正如图 3 所示的，本研究的价值还在于发现配菜规律，发现居民的饮食习惯。生姜、蒜米、小红椒属于厨房必备配料，居民在采购时往往同时下单，关联度之高不言而喻，表明模型是有效的。下面来分析获得的几条有价值的菜品配置推荐规则：

(1) “提升度”最大的是“芹菜－牛肉丝”规则，在衡阳的饭馆或家里餐桌，芹菜炒牛肉味道可口、香味浓是常见菜。尽管菜的做法不一样，“芹菜－牛肉”不会改变菜品色香味的本质，显然切好的牛肉丝更受欢迎。

(2) “白豆腐－雄鱼头”这条规则的提升度为 3.13，也是衡阳居民的餐饮特色之一，“煮鱼头”配豆腐，营养丰富，味道鲜美，常见于餐桌。

(3) “小红椒－老母鸡丁”规则的提升度为 3.416，事实也确实如此，辣椒炒鸡丁是湖南地区居民的特色菜之一，辣与鸡同在，还能上品质。

(4) “瘦肉丝－红萝卜片”规则的提升度为 3.366，反映居民健康的生活习惯。红萝卜炒瘦肉丝，既能补充蛋白质，又能补充维生素，清淡又富营养。

(5) “芹菜－大蒜叶”二者同时配置的方式值得推荐，二者持久的、浓烈的香味、口感放在一起，而且芹菜的降血脂功效让人放心，各地民众都有同感。

5. 结论与展望

本文应用 Apriori 算法来研究连锁净菜门店的菜品配置推荐规则，从 177 种菜品(包含配料)中，在支持度 3%的阈值下，获得了 249 条搭配销售规则，24 条两项规则，具有高推荐价值，规则提升度大于 3，达到了推荐应用效果。IBM modeler 工具封装了 Apriori 模型，使得挖掘工作相对傻瓜化。但是，从获得的菜品推荐规则来看，多条规则与本地居民饮食习惯一致，说明了研究模型的科学性、有效性，也说明其他规则可以应用于实践。另一方面，从已提取的菜品搭配规则来看，无疑能够帮助洗净菜配送中心用来指导编排配送计划，并知道到门店的上架工作。这种搭配实则在引导居民的采购行为，从而加速流通。在进一步的研究中，将在菜品中增加菜品的单次采购量，并进一步研究菜品搭配的比例结构，从而为门店的菜品采购配置提供决策方法，以优化门店库存，提高洗净菜配送中心的加工效益。

参考文献

[1] 薄治禹. 饮食行业数据库中关联模式的卡方分析[J]. 中国商贸, 2013(9): 76-77.

[2] 王楚瑜, 汪庆华, 王楚楚, 陈艳辉, 华祎恒. 高校食堂的菜品销量分析[J]. 计算机时代, 2017(7): 65-68+71.

[3] 李筠, 岳勤霏, 范欣生. 方后注服药食忌研究[J]. 中医杂志, 2018, 59(10): 833-836.

[4] 邱京伟. 订餐系统推荐模块设计[J]. 信息与电脑(理论版), 2018(22): 115-117.

[5] 张奥多, 张昕, 李怡婷. 基于关联规则的餐饮服务智能推荐系统[J]. 广西科技大学学报, 2017, 28(3): 117-123+131.

[6] 谢奇爱, 董宜文. 基于 LBS 的个性化手机菜品推荐系统设计与实现[J]. 重庆科技学院学报(自然科学版), 2017, 19(6): 117-119+124.