

基于关联规则分析的 3 个卷烟价位段 群体消费习惯特征研究

何建龙, 张 涛, 王海娟, 浦 倩, 徐艳群

(云南中烟工业有限责任公司技术中心, 云南 昆明 650231)

摘要:为进一步了解不同卷烟价位段人群的消费习惯特征, 研究对 16 个城市中消费习惯基础调查项目, “好烟”意识判断, 未来关注物理特性、新品类方向和口味新方向进行调研并与 3 个卷烟价位段基于 Apriori 算法进行关联规则分析。结果表明: 具有关联较多的有效规则数主要出现在购买卷烟价位段在 20 ~ 39.9 元的群体上, 而只有 1 条有效规则数出现在购买卷烟价位段在 40 元以上的群体上。结果可为工业在新产品开发方向提供参考价值。

关 键 词:消费习惯; 关联规则分析; 有效规则数

0 引言

关联规则挖掘是从大量的数据中或对象间抽取关联性。揭示数据间未知的依赖关系, 根据这种关联性从某一数据对象的信息来推断另一数据对象的信息。关联规则挖掘是数据挖掘技术中的一个重要课题^[1]。目前关联规则分析在卷烟市场调研上已经有了一定程度的研究: 刘静等人采用关联规则等方法完成了中国卷烟市场调查报告^[2], 黄智等人采用关联规则方法完成了各消费群体对某产品 5 元定价的态度以及该定价是否合理进行分析^[1], 周欣胤等人也利用关联规则 Apriori 算法, 对四川省卷烟市场数据进行研究获得提高卷烟厂销量的指导性意见^[3]。但是以上研究调查的内容属于消费者的背景信息与消费者月均消费进行关联, 主要挖掘了月均消费较高群体的背景特征, 对于工业在新产品开发方向的指导上依然存在空白, 因此本研究对 2015 年 16 个城市的消费习惯进行调研, 从基础调查项目, “好烟”意识判断, 未来关注物理特性、新品类方向和口味新方向与 3 个卷烟价位段进行关联, 从而获得 3 个卷烟价位段消费群体的消费习惯特征, 以期工业在新产品开发方向提供参考价值。

1 材料与方法

1.1 材料

研究针对 2015 年 16 个城市的消费习惯进行调研, 共收集有效样本 1700 条, 选取了基础调查项目, “好烟”意识判断, 未来关注物理特性、新品类方向和口味新方向作为关联规则的前项, 以卷烟价位段: 9 ~

19.9 元(低)、20 ~ 39.9 元(中)、40 元以上(高), 作为关联规则的后项(目标项)进行分析。

基于基础调查项目进行的 11 项变量设置为: (1) 香气: 清香(1st)、浓香(2nd)、中间香(3rd)、未关注(4th)。(2) 劲头: 大(1st)、小(2nd)、适中(3rd)、未关注(4th)。(3) 余味: 余香明显(1st)、无余味(2nd)、有一定余味(3rd)、未关注(4th)。(4) 烟气: 浓(1st)、淡(2nd)、一般(3rd)、未关注(4th)。(5) 香精: 添加多(1st)、添加少(2nd)、无添加(3rd)、未关注(4th)。(6) 吸味: 平淡/醇和(1st)、浓郁/重(2nd)、未关注(3rd)。(7) 包装风格: 怀旧经典(1st)、年轻时尚(2nd)、高贵典雅(3rd)、现代简约(4th)、民族风情(5th)、传统中国风(6th)。(8) 包装颜色: 红色(1st)、黄(金)色(2nd)、蓝色(3rd)、绿色(4th)、白色(5th)、棕色(6th)、银色(7th)、黑色(8th)、无所谓(9th)、其他颜色(10th)。(9) 常吸焦油量: [10 ~ 12mg(1st)、8 ~ 9mg(2nd)、6 ~ 7mg(3rd)、5mg 以下(4th)、未关注(5th)。(10) 卷烟消费月支出: 100 元以下(1st)、101 ~ 200 元(2nd)、201 ~ 300 元(3rd)、301 ~ 400 元(4th)、401 ~ 500 元(5th)、501 ~ 600 元(6th)、601 ~ 700 元(7th)、701 ~ 800 元(8th)、801 ~ 900 元(9th)、901 ~ 1000 元(10th)、1000 ~ 2000 元(11th)、2000 元以上(12th)。(11) 品牌忠诚度: 总是买一个牌子的一种烟(1st)、只买一个牌子, 包括这个牌子下面不同价格的烟(2nd)、相对固定在 2 ~ 3 个牌子的烟上(3rd)、在一段时间内会固定买一种烟, 但过上一阵就会换(4th)、不看牌子, 只要价格合适, 有什么烟就买什么烟(5th)。

基于“好烟”意识判断进行的 11 项变量设置为: (1) 烟丝油亮: 关注(1st)、未关注(2nd)。(2) 烟丝金黄: 关注(1st)、未关注(2nd)。(3) 烟丝长: 关注(1st)、未关注(2nd)。(4) 烟丝填充的均匀程度: 关注(1st)、未关注(2nd)。(5) 过滤嘴有镂空: 关注(1st)、未关注(2nd)。(6) 过滤嘴更长一些: 关注(1st)、未关

2.1.1 9~19.9元价位消费者群体消费习惯中基础调查项目关联规则分析

后项{价位段=低}即购买卷烟价位段在9~19.9元的有效规则,如表6所示。

规则1、4、11说明:卷烟消费月支出为101~400元的消费者人数占总调查人数的24%,其中有超过半数(约58%)的人属于购买卷烟价位段在9~19.9元的群体。

规则7说明:受到清香香气所吸引的消费者人数

占总调查人数的6%,其中有近半数(41%)的人属于购买卷烟价位段在9~19.9元的群体。

规则9说明:常吸卷烟焦油量为10~12mg的消费者人数占总调查人数的6%。其中有近半数(41%)的人属于购买卷烟价位段在9~19.9元的群体。

有效性上,这5条规则的支持度均大于最小阈值0.05,为强规则。在实用性上,这5条规则的提升度远大于1,即前项对后项影响非常大。所以给出的规则既是有效的又是实用。

表6 后项{价位段=低}的有效规则

有效规则	前项	后项	支持度	置信度	提升度
1	{卷烟消费月支出=3rd}	{价位段=低}	0.1	0.65	2.08
4	{卷烟消费月支出=4th}	{价位段=低}	0.08	0.41	1.3
7	{香气=1st,余味=4th,烟气=4th,香精=4th}	{价位段=低}	0.06	0.41	1.31
9	{烟气=4th,吸味=3rd,常吸焦油量=1st}	{价位段=低}	0.06	0.41	1.3
11	{卷烟消费月支出=2nd}	{价位段=低}	0.06	0.68	2.18

2.1.2 20~39.9元价位消费者群体消费习惯中基础调查项目关联规则分析

后项{价位段=中}即购买卷烟价位段在20~39.9元的有效规则,如表7所示。

规则6、24、42、60说明:买烟时具有一定忠诚度(相对固定在2~3个牌子的烟上、买烟时只买一个牌子,包括这个牌子下面不同价格的烟)且对劲头要求适中的消费者人数占总调查人数的37%,其中有超过(51%)半数的人属于购买卷烟价位段在20~39.9元的群体;买烟时具有一定忠诚度较低(买烟时在一段时间内会固定买一种烟,但过上一阵就会换)的消费者人数占总调查人数的6%,其中有超过(59%)半数的人属于购买卷烟价位段在20~39.9元的群体。

规则8、78说明:最喜欢的香烟外观包装主体颜色为暖色系(红色、黄(金)色)的消费者人数占总调查人数的20%,其中有超过(51%)半数的人属于购买卷烟价位段在20~39.9元的群体。

规则9说明:常吸卷烟焦油量为10~12mg的消费者人数占总调查人数的15%,其中有超过(50%)半数的人属于购买卷烟价位段在20~39.9元的群体。

规则14、20说明:最喜欢的香烟外观包装风格是年轻时尚的消费者人数占总调查人数的12%,最喜欢的香烟外观包装风格是高贵典雅的消费者人数占总调查人数的10%,其中两者均有超过(52%)半数的人属于购买卷烟价位段在20~39.9元的群体。

规则18、35、43、48、59说明:卷烟消费月支出为

301~800元的消费者人数占总调查人数的37%。其中有超过约(63%)半数的人属于购买卷烟价位段在20~39.9元的群体。

规则19说明:受到吸味平淡/醇和所吸引的消费者人数占总调查人数的10%。其中有超过(52%)半数的人属于购买卷烟价位段在20~39.9元的群体。

规则26、37说明:受到清香香气所吸引的消费者人数占总调查人数的9%。受到中间香香气所吸引的消费者人数占总调查人数的7%。其中两者均有超过(52%)半数的人属于购买卷烟价位段在20~39.9元的群体。

规则34说明:受到有一定余味所吸引的消费者人数占总调查人数的7%。其中有近(50%)半数的人属于购买卷烟价位段在20~39.9元的群体。

规则47、74说明:受到香精添加少或香精无添加所吸引的消费者人数占总调查人数的11%。其中有超过(约56%)半数的人属于购买卷烟价位段在20~39.9元的群体。

规则61说明:受到烟气淡所吸引的消费者人数占总调查人数的6%。其中有超过(62%)半数的人属于购买卷烟价位段在20~39.9元的群体。

有效性上,这21条规则的支持度均大于最小阈值0.05,为强规则。在实用性上,这21条规则的提升度远大于1,即前项对后项影响非常大。所以给出的规则既是有效的又是实用。

表 7 后项|价位段=中|的有效规则

有效规则	前项	后项	支持度	置信度	提升度
6	{ 品牌忠诚度 = 3rd }	{ 价位段 = 中 }	0.21	0.51	1.03
8	{ 包装颜色 = 1st }	{ 价位段 = 中 }	0.15	0.50	1.01
9	{ 常吸焦油量 = 1st }	{ 价位段 = 中 }	0.15	0.50	1.01
14	{ 包装风格 = 2nd }	{ 价位段 = 中 }	0.12	0.53	1.06
18	{ 卷烟消费月支出 = 4th }	{ 价位段 = 中 }	0.11	0.57	1.15
19	{ 余味 = 4th, 吸味 = 1st }	{ 价位段 = 中 }	0.10	0.52	1.05
20	{ 包装风格 = 3rd }	{ 价位段 = 中 }	0.10	0.51	1.03
24	{ 劲头 = 3rd, 品牌忠诚度 = 3rd }	{ 价位段 = 中 }	0.09	0.52	1.05
26	{ 香气 = 1st, 劲头 = 4th }	{ 价位段 = 中 }	0.09	0.54	1.08
34	{ 余味 = 3rd, 香精 = 4th }	{ 价位段 = 中 }	0.07	0.50	1.01
35	{ 卷烟消费月支出 = 7th }	{ 价位段 = 中 }	0.07	0.75	1.50
37	{ 香气 = 3rd, 余味 = 4th }	{ 价位段 = 中 }	0.07	0.50	1.01
42	{ 品牌忠诚度 = 2nd }	{ 价位段 = 中 }	0.07	0.50	1.01
43	{ 卷烟消费月支出 = 6th }	{ 价位段 = 中 }	0.07	0.67	1.35
47	{ 香精 = 2nd }	{ 价位段 = 中 }	0.06	0.54	1.08
48	{ 卷烟消费月支出 = 5th }	{ 价位段 = 中 }	0.06	0.51	1.03
59	{ 卷烟消费月支出 = 8th }	{ 价位段 = 中 }	0.06	0.69	1.39
60	{ 品牌忠诚度 = 4th }	{ 价位段 = 中 }	0.06	0.59	1.19
61	{ 烟气 = 2nd }	{ 价位段 = 中 }	0.06	0.62	1.25
74	{ 香精 = 3rd }	{ 价位段 = 中 }	0.05	0.65	1.31
78	{ 劲头 = 4th, 包装颜色 = 2nd }	{ 价位段 = 中 }	0.05	0.53	1.06

2.1.3 40 元以上价位消费者群体消费习惯中基础调查项目关联规则分析

后项|价位段=高|即购买卷烟价位段在 40 元以上的有效规则,如表 8 所示。

规则 6 说明:卷烟消费月支出为 1000 ~ 2000 元的消费者人数占总调查人数的 6%。其中有超过(77%)

半数的人属于购买卷烟价位段在 40 元以上的群体。有效性上,这 1 条规则的支持度均大于最小阈值 0.05,为强规则。在实用性上,这 1 条规则的提升度远大于 1,即前项对后项影响非常大。所以给出的规则既是有效的又是实用。

表 8 后项|价位段=高|的有效规则

有效规则	前项	后项	支持度	置信度	提升度
6	{ 卷烟消费月支出 = 11th }	{ 价位段 = 高 }	0.06	0.77	4.03

2.2 3 个价位消费者群体消费习惯中“好烟”意识判断关联规则分析

通过对选取中国卷烟消费习惯中“好烟”意识判断的 11 个观测指标作为关联规则的前项,以卷烟价位段作为关联规则的后项,参照以往的类似研究参数设定,研究设定最小支持度为 10%,最小置信度为 40%,通过 Apriori 算法进行分析,筛选提升度为 1 以上且支持度较高的有效规则(即选项有参考意义)。分析结果发现:没有找到与 9 ~ 19.9 元卷烟价位段群体具有关联的有效规则,与 20 ~ 39.9 元卷烟价位段群体具有关联的有效规则数为 4 条,没有找到与 40 元以上卷烟价位段群体具有关联的有效规则。图 4 显示了出现最频繁的频率在 10% 以上的项集。

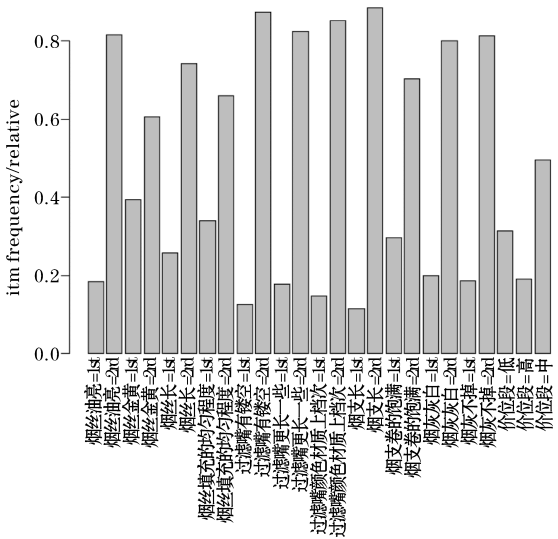


图 4 在“好烟”意识判断方面出现频率在 10% 以上的项集频率

2.2.1 20~39.9元价位消费者群体消费习惯中“好烟”意识判断关联规则分析

后项{价位段=中}即购买卷烟价位段在20~39.9元的有效规则,如表9所示。

规则65、72、102说明:判断烟的好坏基于烟丝金黄的消费者人数占总调查人数的16%。判断烟的好坏基于烟丝填充的均匀程度的消费者人数占总调查人数的15%。判断烟的好坏基于烟丝长的消费者人数占总调查人数的13%。其中三者均有超过(50%)半

数的人属于购买卷烟价位段在20~39.9元的群体。

规则160说明:判断烟的好坏基于烟灰灰白的消费者人数占总调查人数的11%。其中有超过(55%)半数的人属于购买卷烟价位段在20~39.9元的群体。

有效性上,这4条规则的支持度均大于最小阈值0.1,为强规则。在实用性上,这4条规则的提升度远大于1,即前项对后项影响非常大。所以给出的规则既是有效的又是实用。

表9 后项{价位段=中}的有效规则

有效规则	前项	后项	支持度	置信度	提升度
65	{烟丝油亮=2nd,烟丝金黄=1st}	{价位段=中}	0.16	0.50	1.01
72	{烟丝油亮=2nd,烟丝填充的均匀程度=1st}	{价位段=中}	0.15	0.51	1.03
102	{烟丝长=1st}	{价位段=中}	0.13	0.51	1.03
160	{烟灰灰白=1st}	{价位段=中}	0.11	0.55	1.10

2.3 3个价位消费者群体消费习惯中未来关注物理特性关联规则分析

通过对选取中国卷烟消费习惯中未来关注物理特性的9个观测指标作为关联规则的前项,以卷烟价位段作为关联规则的后项,参照以往的类似研究参数设定,研究设定最小支持度为5%,最小置信度为40%,通过Apriori算法进行分析,筛选提升度为1以上且支持度较高的有效规则(即选项有参考意义)。分析结果发现:与9~19.9元卷烟价位段群体具有关联的有效规则数为1条,与20~39.9元卷烟价位段群体具有关联的有效规则数为6条,没有找到与40元以上卷烟价位段群体具有关联的有效规则。图5显示了出现最频繁的频率在5%以上的项集。

2.3.1 9~19.9元价位消费者群体消费习惯中未来关注物理特性关联规则分析

后项{价位段=低}即购买卷烟价位段在9~19.9元的有效规则,如表10所示。

规则1说明:未来会关注细支烟的消费者人数占总调查人数的5%。其中有接近(41%)半数的人属于购买卷烟价位段在9~19.9元的群体。

有效性上,这1条规则的支持度均大于最小阈值0.05,为强规则。在实用性上,这1条规则的提升度远大于1,即前项对后项影响非常大。所以给出的规则既是有效的又是实用。

2.3.2 20~39.9元价位消费者群体消费习惯中未来关注物理特性关联规则分析

后项{价位段=中}即购买卷烟价位段在20~39.9元的有效规则,如表11所示。

规则15、34说明:未来会关注烟支比一般长一些的消费者人数占总调查人数的10%。未来会关注烟支比一般短一些的消费者人数占总调查人数的7%。其中两者均有超过(52%)半数的人属于购买卷烟价位段在20~39.9元的群体。

规则24、32、33说明:未来会关注过滤嘴镂空的消费者人数占总调查人数的8%。未来会关注过滤嘴同轴滤棒的消费者人数占总调查人数的7%。未来会关注过滤嘴特色——爆珠的消费者人数占总调查人数的7%。其中三者均有超过(52%)半数的人属于购买卷烟价位段在20~39.9元的群体。

规则25说明:未来会关注混合型香烟的消费者人数占总调查人数的8%。其中有超过(52%)半数的人属于购买卷烟价位段在20~39.9元的群体。

有效性上,这6条规则的支持度均大于最小阈值

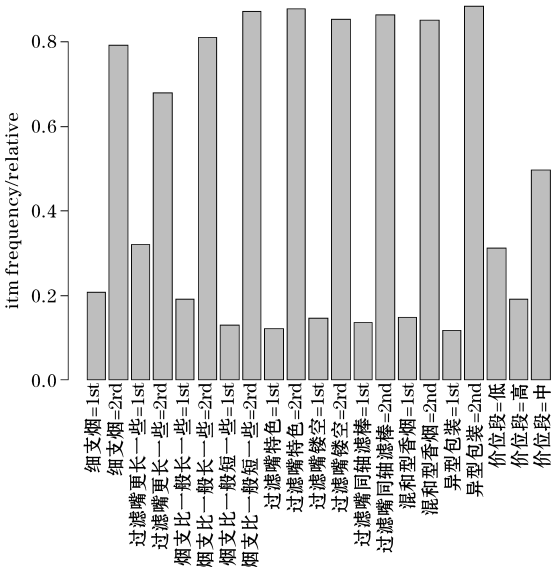


图5 在未来关注物理特性方面出现频率在5%以上的项集频率

0.05,为强规则。在实用性上,这 6 条规则的提升度远 是有效的又是实用。
大于 1,即前项对后项影响非常大。所以给出的规则既

表 10 后项{价位段=低}的有效规则

有效规则	前项	后项	支持度	置信度	提升度
1	{细支烟=1st,过滤嘴更长一些=2nd,烟支比一般长一些=2nd,混和型香烟=2nd,异型包装=2nd}	{价位段=低}	0.05	0.41	1.3

表 11 后项{价位段=中}的有效规则

有效规则	前项	后项	支持度	置信度	提升度
15	{烟支比一般长一些(如 94mm)=1st}	{价位段=中}	0.10	0.53	1.07
24	{过滤嘴镂空=1st}	{价位段=中}	0.08	0.54	1.10
25	{混和型香烟(如外烟、中南海等)=1st}	{价位段=中}	0.08	0.52	1.05
32	{过滤嘴同轴滤棒(过滤焦油含量可调整)=1st}	{价位段=中}	0.07	0.50	1.01
33	{过滤嘴特色——爆珠的=1st}	{价位段=中}	0.07	0.55	1.11
34	{烟支比一般短一些(如黄鹤楼 1916)=1st}	{价位段=中}	0.07	0.52	1.05

2.4 3 个价位消费者群体消费习惯中未来关注新品类方向关联规则分析

通过对选取中国卷烟消费习惯中未来关注新品类的 4 个观测指标作为关联规则的前项,以卷烟价位段作为关联规则的后项,参照以往的类似研究参数设定,研究设定最小支持度为 5%,最小置信度为 40%,通过 Apriori 算法进行分析,筛选提升度为 1 以上且支持度较高的有效规则(即选项有参考意义)。分析结果发现:没有找到 9~19.9 元卷烟价位段群体具有关联的有效规则,与 20~39.9 元卷烟价位段群体具有关联的有效规则数为 1 条,没有找到与 40 元以上卷烟价位段群体具有关联的有效规则。图 6 显示了出现最频繁的频率在 5% 以上的项集。

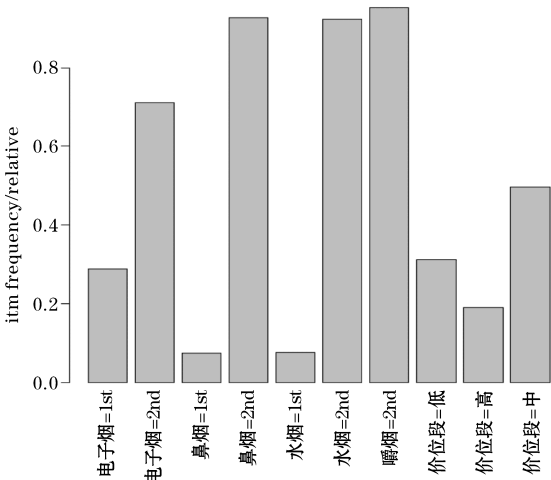


图 6 在未来关注新品类方向方面出现频率在 5% 以上的项集频率

表 12 后项{价位段=中}的有效规则

有效规则	前项	后项	支持度	置信度	提升度
9	{电子烟=1st}	{价位段=中}	0.14	0.50	1.00

2.4.1 20~39.9 元价位消费者群体消费习惯中未来关注新品类方向关联规则分析

后项{价位段=中}即购买卷烟价位段在 20~39.9 元的有效规则,如表 12 所示。
规则 9 说明:未来会关注电子烟的消费者人数占总调查人数的 14%。其中有超过(50%)半数的人属于购买卷烟价位段在 20~39.9 元的群体。
有效性上,这 1 条规则的支持度均大于最小阈值 0.05,为强规则。在实用性上,这 1 条规则的提升度远大于 1,即前项对后项影响非常大。所以给出的规则既是有效的又是实用。

2.5 3 个价位消费者群体消费习惯中未来关注口味新方向关联规则分析

通过对选取中国卷烟消费习惯中未来关注口味新方向的 8 个观测指标作为关联规则的前项,以卷烟价位段作为关联规则的后项,参照以往的类似研究参数设定,研究设定最小支持度为 5%,最小置信度为 40%,通过 Apriori 算法进行分析,筛选提升度为 1 以上且支持度较高的有效规则(即选项有参考意义)。分析结果发现:没有找到 9~19.9 元卷烟价位段群体具有关联的有效规则,与 20~39.9 元卷烟价位段群体具有关联的有效规则数为 3 条,没有找到与 40 元以上卷烟价位段群体具有关联的有效规则。图 7 显示了出

现最频繁的频率在5%以上的项集。

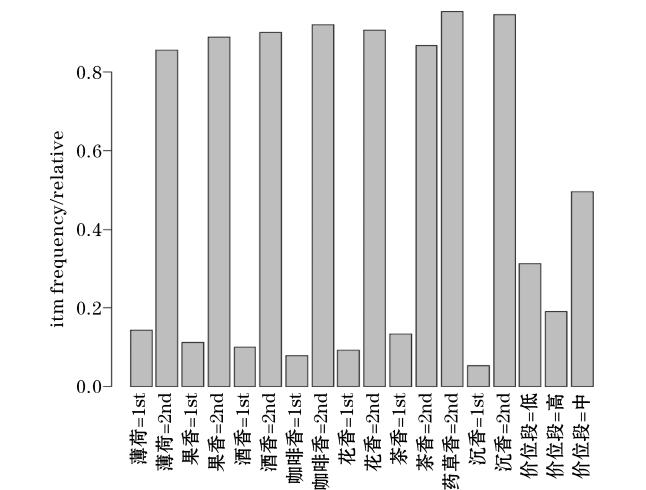


图7 在未来关注口味新方向方面出现频率在5%以上的项集频率

2.5.1 20~39.9元价位消费者群体消费习惯中未来关注口味新方向关联规则分析

后项{价位段=中}即购买卷烟价位段在20~39.9元的有效规则,如表13所示。

规则25、27、45说明:未来会关注茶香的消费者人数占总调查人数的7%。未来会关注果香的消费者人数占总调查人数的6%。未来会关注花香的消费者人数占总调查人数的5%。其中有超过(52%)半数的人属于购买卷烟价位段在20~39.9元的群体。

有效性上,这3条规则的支持度均大于最小阈值0.05,为强规则。在实用性上,这3条规则的提升度远大于1,即前项对后项影响非常大。所以给出的规则既是有效的又是实用。

表13 后项{价位段=中}的有效规则

有效规则	前项	后项	支持度	置信度	提升度
25	{咖啡香=2nd,茶香=1st}	{价位段=中}	0.07	0.53	1.07
27	{果香=1st,沉香=2nd}	{价位段=中}	0.06	0.51	1.02
45	{花香=1st}	{价位段=中}	0.05	0.54	1.08

3 结论与讨论

综合消费习惯中的基础调查项目,“好烟”意识判断以及未来关注的物理特性、新品类方向、口味新方向的关联规则分析总体情况看,具有关联较多的有效规则数主要出现在购买卷烟价位段在20~39.9元的群体上,而只有1条有效规则数出现在购买卷烟价位段在40元以上的群体上。

购买卷烟价位段在9~19.9元即低价位段的群体特征通常表现为:卷烟消费月支出为101~300元,吸味方面主要受到清香香气所吸引,常吸卷烟焦油量为10~12mg,未来会对细支烟类型产品进行关注。

购买卷烟价位段在20~39.9元即中价位段的群体特征通常表现为:在品牌忠诚度方面,大多数相对固定在2~3个牌子的烟上,一部人会只买一个牌子,包括这个牌子下面不同价格烟或是在一段时间内会固定买一种烟,但过上一阵就会换,最喜欢的香烟外观包装主体颜色为红色和黄(金)色的暖色系元素,常吸卷烟焦油量为10~12mg,最喜欢的香烟外观包装风格是年轻时尚和高贵典雅,卷烟消费月支出为301~800元,在吸味方面主要受到吸味平淡/醇和、烟气淡、劲头适中、有一定余味、香精添加少或无添加、清香香气和中间香香气所吸引,判断烟的好坏主要基于烟丝金黄、烟丝填充的均匀程度、烟丝长以及烟灰灰白,未来会对

烟支比一般长一些或短一些、过滤嘴镂空、混合型香烟、过滤嘴同轴滤棒、过滤嘴特色——爆珠的、电子烟以及口味为茶香、果香和花香类型产品进行关注。

购买卷烟价位段在40元以上即高价位段的群体特征通常表现为:卷烟消费月支出为1000~2000元。

参考文献:

[1] 黄智,何跃. 关联规则挖掘在卷烟市场研究中的应用[J]. 成都信息工程学院学报,2005,20(3): 319-323.

[2] 刘静. 中国卷烟市场调查研究报告[D]. 昆明: 云南财经大学,2013.

[3] 周欣胤,何跃. 基于Apriori算法的四川省卷烟市场的研究[J]. 中国商贸,2010(4):13-14.

[4] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases[J]. In Proceedings of the ACM SIGMOD International Conference on Management of Data, 1993:207-216.