

**目录**

[一、主题分类 1](#_Toc39177655)

[1.1风险（risk\_module） 1](#_Toc39177656)

[1.2投资（investment\_module） 2](#_Toc39177657)

[1.3知识产权（creativity\_module） 3](#_Toc39177658)

[1.4品牌（brand\_module） 4](#_Toc39177659)

[1.5招聘（recruit\_module） 4](#_Toc39177660)

[1.6信用（credit\_module） 4](#_Toc39177661)

[1.7企业基本信息（company\_baseinfo\_module） 5](#_Toc39177662)

[二、属性预处理 5](#_Toc39177663)

[2.1风险预处理 6](#_Toc39177664)

[2.2投资预处理 7](#_Toc39177665)

[2.3知识产权预处理 8](#_Toc39177666)

[2.4品牌预处理 8](#_Toc39177667)

[2.5招聘预处理 8](#_Toc39177668)

[2.6信用预处理 8](#_Toc39177669)

[2.7企业基本信息预处理 9](#_Toc39177670)

[三、模块统计 9](#_Toc39177671)

[3.1风险统计 9](#_Toc39177672)

[3.2投资统计 10](#_Toc39177673)

[3.3知识产权统计 11](#_Toc39177674)

[3.4品牌统计 11](#_Toc39177675)

[3.5招聘统计 12](#_Toc39177676)

[3.6信用统计 12](#_Toc39177677)

[3.7企业基本信息统计 12](#_Toc39177678)

[四、模块总评 13](#_Toc39177679)

## 主题分类

由于命题方提供的数据种类丰富，为了对其更好的归类和集中处理，所以首先根据数据所描述的类型，分成了下述的7个模块：风险、投资、知识产权（创新能力）、品牌、招聘、信用、基本信息，并描述了其数据库的建立标准。

其中对于每张表，有以下说明事项：

* 所有的表都有属性id，类型为int，自增主键作为该表的唯一标识，并使用聚簇索引。
* 所有的表都有属性entname（企业），类型为varchar(255)，not null，在下面基本表的描述中不再赘述。
* 日期格式全部统一为YYYY/MM/DD格式，如果是UNIX时间戳的需要提前转化后导入。

### 1.1风险（risk\_module）

表1-1所示是企业风险模块中每张表以及其对应属性的解释。

表1-1：risk\_module基本表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 |
| administrative  \_punishment | is\_punish | int | 默认值0 | 公司行政处罚次数 |
| business\_risk  \_abnormal | is\_bra | int | 默认值0 | 列入经营异常次数 |
| business\_risk  \_all\_punish | is\_brap | int | 默认值0 | 行政处罚记录次数 |
| business\_risk  \_taxunpaid | taxunpaidnum | double | 默认值0.0 | 企业累计欠税额 |
| business\_risk\_rightpledge | pledgenum | int | 默认值0 | 企业股权出质次数 |
| ent\_social  \_security | unpaidsocialins  \_so110 | double | 默认值0.0 | 单位参加城镇职工基本养老保险累计欠缴金额 |
| unpaidsocialins  \_so210 | double | 默认值0.0 | 单位参加失业保险累计欠缴金额 |
| unpaidsocialins  \_so310 | double | 默认值0.0 | 单位参加职工基本医疗保险累计欠缴金额 |
| unpaidsocialins  \_so410 | double | 默认值0.0 | 单位参加工伤保险累计欠缴金额 |
| unpaidsocialins  \_so510 | double | 默认值0.0 | 单位参加生育保险累计欠缴金额 |
| updatetime | date | 可为空 | 更新日期 |
| exception\_list | is\_except | int | 默认值0 | 异常次数 |
| justice\_declare | declaredate | date | YYYY/MM/DD | 公告时间 |
| appellant | int | 默认值0 | 上诉方(企业如果为上诉方，值为1，否则值为0） |
| defendant | int | 默认值0 | 被诉方（企业如果为被诉方，值为1，否则值为0） |
| declarestyle | varchar(255) | 可为空 | 公告类型 |
| justice  \_enforced | record\_date | date | YYYY/MM/DD | 执行日期 |
| enforce\_amount | double | 默认值0.0 | 执行标的金额 |
| justice\_judge  \_new | time | varchar(255) | 可为空 | 时间 |
| title | varchar(255) | 可为空 | 标题 |
| casetype | varchar(255) | 可为空 | 案件类型 |
| judgeresult | text | 可为空 | 判决结果，字符较多 |
| casecause | varchar(255) | 可为空 | 案由 |
| evidence | varchar(255) | 可为空 | 案由编码类型 |
| courtrank | varchar(255) | 可为空 | 依据 |
| datatype | varchar(255) | 可为空 | 法院等级 |
| latypes | varchar(255) | 可为空 | 司法类型 |
| justice\_credit | is\_justice\_credit | int | 默认值0 | 是否列入失信黑名单 |
| justice\_credit  \_aic | is\_justice\_creditaic | int | 默认值0 | 是否工商部失信企业 |

### 1.2投资（investment\_module）

表1-2所示是企业投资模块每张表以及其对应属性的解释。

表1-2：investment\_module基本表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 |
| ent\_bid | bidnum | int | 默认值0 | 中标次数 |
| ent\_branch | branchnum | int | 默认值0 | 企业分支数 |
| ent\_contribution | invtype | varchar(255) | 可为空 | 投资人类型 |
| conform | varchar(255) | 可为空 | 出资方式 |
| subconam | double | 默认值0.0 | 认缴出资额 |
| conprop | int | 默认值0 | 持股比例 |
| condate | date | 可为空 | 出资日期 |
| ent\_contribution\_year | subconcurrency | varchar(255) | 可为空 | 认缴币种 |
| accondate | date | 可为空 | 实缴出资时间 |
| subconform | varchar(255) | 可为空 | 认缴出资方式 |
| anchetype | varchar(255) | 可为空 | 行业分类 |
| subcondate | date | 可为空 | 认缴出资时间 |
| acconcurrency | varchar(255) | 可为空 | 实缴币种 |
| acconform | varchar(255) | 可为空 | 实缴出资方式 |
| liacconam | double | 默认值0.0 | 累计实缴额 |
| lisubconam | double | 默认值0.0 | 累计认缴额 |
| ent\_guarantee | priclaseckind | double | 默认值0.0 | 主债权种类 |
| pefperfrom | date | 可为空 | 履行债务的期限自 |
| iftopub | varchar(255) | 可为空 | 是否公示此担保信息1是2否 |
| priclasecam | double | 默认值0.0 | 主债权数额 |
| pefperto | date | 可为空 | 履行债务的期限至 |
| guaranperiod | int | 默认值0 | 保证的期间1期限2未约定 |
| gatype | int | 默认值0 | 保证的方式1一般保证2连带保证3未约定 |
| rage | varchar(255) | 默认值“0” | 保证担保的范围 |
| ent\_investment | investnum | int | 默认值0 | 投资次数 |
| ent\_onlineshop | shopnum | int | 默认值0 | 网店个数 |
| enterprise  \_insurance | cbrq | date | 可为空 | 参保日期 |
| xzbz | varchar(255) | 可为空 | 险种标志 |
| sbjgbh | varchar(255) | 可为空 | 社会保险经办机构 |
| xzbzmc | varchar(255) | 可为空 | 险种标志名称 |
| cbzt | varchar(255) | 可为空 | 参保状态 |
| cbztmc | varchar(255) | 可为空 | 参保状态名称 |
| dwbh | varchar(255) | 可为空 | 单位编号 |

### 1.3知识产权（creativity\_module）

表1-3所示是知识产权模块每张表以及其对应属性的解释。

表1-3：creativity\_module基本表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 |
| intangible  \_brand | ibrand\_num | int | 默认值0 | 知识产权--商标申请次数 |
| intangible  \_copyright | icopy\_num | int | 默认值0 | 企业软件著作权登记次数 |
| intangible  \_patent | ipat\_num | int | 默认值0 | 企业专利申请次数 |
| web\_record\_info | idom\_num | int | 默认值0 | 企业是否拥有域名的知识产权 |

### 1.4品牌（brand\_module）

表1-4所示是品牌模块每张表以及其对应属性的解释。

表1-4：brand\_module基本表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 |
| jn\_special  \_new\_info | is\_jnsn | int | 默认值0 | 是否是济南市专精特新中小企业（缺失值99.99%） |
| jn\_tech\_center | level\_rank | int | 默认值0 | 级别(省级2、市级1、企业名称不出现在该表则值为0) |
| trademark\_infoa | is\_infoa | int | 默认值0 | 是否列为驰名商标 |
| trademark\_infob | is\_infob | int | 默认值0 | 是否列为著名商标 |
| product  \_checkinfo  \_connect | passpercent | double | 默认值0.0 | 企业产品被抽查的合格率（未被抽查值为0，被抽查则值为0-1之间小数值） |

### 1.5招聘（recruit\_module）

表1-5所示是企业招聘模块每张表以及其对应属性的解释。

表1-5：recruit\_module基本表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 |
| recruit\_module | qcwynum | int | 默认值0 | 前程无忧招聘数 |
| zhycnum | int | 默认值0 | 中华英才招聘数 |
| zlzpnum | int | 默认值0 | 智联招聘招聘数 |

### 1.6信用（credit\_module）

表1-6所示是企业信用模块每张表以及其对应属性的解释。

表1-6：credit\_module基本表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 |
| enterprise  \_keep\_contract | is\_kcont | int | 默认值0 | 是否列为守合同重信用企业 |
| jn\_credit\_info | credit\_grade | varchar(255) | 默认值N | 信用等级 N+、B-、A、C、N、A- |

### 1.7企业基本信息（company\_baseinfo\_module）

表1-7所示是企业基本信息模块每张表以及其对应属性的解释。

表1-7：baseinfo\_module基本表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 |
| company  \_baseinfo | regcap | double | 默认值0.0 | 注册资本 |
| empnum | int | 默认值0 | 从业人数 |
| estdate | date | 默认值“0” | 成立日期 |
| candate | date | 可为空 | 注销时间 |
| revdate | date | 可为空 | 吊销时间 |
| entstatus | varchar(255) | 可为空 | 企业状态 |
| opto | date | 可为空 | 经营（驻在）期限至 |
| enttype | varchar(255) | 可为空 | 企业(机构)类型 |
| entcat | varchar(255) | 可为空 | 企业类别 |
| industryphy | varchar(255) | 可为空 | 行业门类 |
| regcapcur | varchar(255) | 可为空 | 注册资本(金)币种 |
| industryco | varchar(255) | 可为空 | 业务类型 |
| opfrom | date | 可为空 | 经营(驻在)期限自 |
| change\_info | remark | varchar(255) | 可为空 | 备注 |
| dataflag | int | 可为空 | 数据来源标志：1核准通过2删除或者驳回或者不予受理 |
| alttime | int | 可为空 | 变更次数 |
| altitem | varchar(255) | 可为空 | 变更事项 |
| cxstatus | int | 可为空 | 撤销状态：  1：变更  2：撤销变更  3：已撤销变更 |
| altdate | date | 可为空 | 变更日期 |
| openo | varchar(255) | 可为空 | 业务编号 |

## 属性预处理

在这一步中，需要对每张表中的属性进行筛选。对于筛选出的属性，默认使用K-Means算法，对每一个属性（一维）进行聚类，聚成5类，并根据其kmeans.center值进行排序，相当于将各种属性统一到5个等级的度量。此外还有以下说明事项：

* 新生成的表默认拥有id和entname属性，要求与“主题分类”相同
* 对于“x”属性用聚类得到的标签。我们用“x\_type”进行标识，并在原表附加该列。
* 对于那些属性值为空或者0的属性，我们采用0来标识。
* 对于“y”表，可能会由处理后，生成新的表，若新表无特殊含义，则用y\_p（process）表示。所有新生成的中间表用蓝色标出。
* 对于只需要用K-Means聚类，而不需要额外描述的表和属性，直接用“K-Means聚类”表示其处理方式，得到x\_type。

在2.1-2.7中，将会根据每个表，筛选出待定的用于建模的属性，并描述它们的处理方式，和新生成的中间表。

### 2.1风险预处理

表2-1所示是风险模块的预处理方式。

表2-1：风险预处理

|  |  |  |
| --- | --- | --- |
| 表名 | 有用属性 | 处理方式 |
| administrative  \_punishment | is\_punish | K-Means聚类 |
| business\_risk  \_abnormal | is\_bra |
| business\_right\_pledge | pledgenum |
| business  \_risk\_all\_punish | is\_brap |
| business\_risk  \_taxunpaid | taxunpaidnum | 一个公司可能存在多条记录，计算每个公司的欠税总和，使用K-Means聚类 |
| ent\_social\_security | unpaidsocialins\_so210-510 | 计算四种保险的欠款总额，得到新表ent\_social\_security\_p，保险欠缴总额用unpaid\_sum(int)表示，并对改属性使用K-Means聚类 |
| exception\_list\_test | is\_except | K-Means聚类 |
| justice\_declare | declaredate，  appellant,  defendant | 上诉方和被告方，统计每个公司上诉和被告的总次数，以及最新的纠纷日期，日期用365\*year+30\*month+day处理，生成一个新表justice\_declare\_p表， declaredate (int),appellant\_amount(int)，  defendant\_amount(int)，并对三个属性进行K-Means聚类 |
| justice\_enforced | record\_date，  enforced\_amount | 统计出每一个公司最近一次的被执行日期（即x=365\*年+30\*月+天的最大值，最终新表显示的值就是x），以及执行金额。生成新表justice\_enforced\_p，分别对record\_date和enforced\_amount进行K-Means聚类。 |
| justice\_judge\_new | / | 统计每个企业司法纠纷的次数，生成新表justice\_judge\_new\_count，  和属性judge\_new\_count（司法纠纷记录次数，int），使用K-Means聚类 |
| justice\_credit | is\_justice\_credit | K-Means聚类 |

### 2.2投资预处理

表2-2所示是投资模块的预处理方式。

表2-2：投资预处理

|  |  |  |
| --- | --- | --- |
| 表名 | 有用属性 | 处理方式 |
| ent\_bid | bidnum | K-Means聚类 |
| ent\_branch | branchnum |
| ent\_contribution | subconam，conprop | 分别是认缴出资额和持股比例。一个公司可能会有多条记录，表示多个出资人，我们希望获得一个公司的总出资额。  ①如果一条公司的多个记录当中，认缴出资额和持股比例均不为0，则总出资额=该条记录认缴出资额/该条记录持股比例；  ②如果持股比例这项的数据都是缺失的，则总出资额为所有记录认缴出资额之和  ③生成新表ent\_contribution\_total  ，属性subconam\_total(double)，并进行K-Means聚类 |
| ent\_contribution\_year | liacconam,lisubconam | 累计实缴额和累计认缴额，同理一个公司可能出现多次，各自求出每个公司累计认缴和累计实缴的总和，然后生成新表ent\_contribution\_year\_total，  属性名就是原来的liacconam,  lisubconam,并用K-Means聚类 |
| ent\_guarantee | / | 不纳入评估 |
| ent\_investment | investnum | K-Means聚类 |
| ent\_onlineshop | shopnum |
| enterprise\_insurance | cbrq | cbrq是参保日期，根据数值规律得到前4位代表的是参保年份。每个企业会有多条记录，表示可能在不同年份参保了不同份数的保险。我们统计每个企业每年参保的保险平均份数。例如某公司有6条2016年的提交保险的记录，3条2013年提交保险的记录，则该企业年平均提交保险份数为4.5份，统一向上取整汇总结果生成新表  enterprise\_insurance\_year\_avg，  属性为insurance\_num\_avg  (double)，并用K-Means聚类。 |

### 2.3知识产权预处理

表2-3所示是知识产权模块的预处理方式。

表2-3：知识产权预处理

|  |  |  |
| --- | --- | --- |
| 表名 | 有用属性 | 处理方式 |
| intangible\_brand | ibrand\_num | K-Means聚类 |
| intangible\_copyright | icopy\_num |
| intangible\_patent | ipat\_num |
| web\_record\_info | idom\_num |

### 2.4品牌预处理

表2-4所示是品牌模块的预处理方式。

表2-4：品牌预处理

|  |  |  |
| --- | --- | --- |
| 表名 | 有用属性 | 处理方式 |
| jn\_special\_new\_info | is\_jnsn | K-Means聚类 |
| jn\_tech\_center | level\_rank |
| trademark\_infoa | is\_infoa |
| trademark\_infob | is\_infob |
| product\_checkinfo\_connect | passpercent |

### 2.5招聘预处理

表2-5所示是知识产权模块的预处理方式。

表2-5：招聘预处理

|  |  |  |
| --- | --- | --- |
| 表名 | 有用属性 | 处理方式 |
| recruit\_module | qcwynum,  zhycnum,  zlzpnum | 三家平台的招聘总数得到recruit\_sum，并做K-Means聚类 |

### 2.6信用预处理

表2-6所示是信用模块的预处理方式。

表2-6：信用预处理

|  |  |  |
| --- | --- | --- |
| 表名 | 有用属性 | 处理方式 |
| enterprise\_keep\_contract | is\_kcont | K-Means聚类 |
| jn\_credit\_info | credit\_grade | credit\_level\_dict = {  'C-':1, 'B-':2,'A-':3,  'A':4, 'N-':5,'N':6  }  将等级映射成对应的分数，使用K-Means聚类。 |

### 2.7企业基本信息预处理

表2-7所示是基本信息模块的预处理方式。

表2-7：企业基本信息预处理

|  |  |  |
| --- | --- | --- |
| 表名 | 有用属性 | 处理方式 |
| company\_base\_info | regcap，empnum，esdate | 不用建立新表，但是要对这三个属性，分别kmeans聚类打分。此外只筛选在营企业，对它进行打分。 |
| change\_info | / | 不纳入评估 |

## 模块统计

在对目标属性进行聚类后，相当于是将范围不同的值，压缩到了同一个范围（1-5），这样才能更好的将多个属性，放在一起综合评估。

根据每个属性对某个模块或者主题的影响程度不同，可以设置不同的权重，进行加权求和，得到某个模块的评价总分。通过这种方式，实际上将企业的多个属性最后压缩到了7个维度，既尽最大可能，避免了维度压缩带来的信息损失，也避免了部分属性缺失，对评估造成的麻烦。

在模块汇总过程中，有以下说明：

* 新生成的表默认拥有id和entname属性，要求与“主题分类”相同。
* 生成模块总评即xxx\_module，即各个module内每个属性type的加权求和值。
* 每个模块总评xxx\_module属性，都使用K-Means聚类，赋予1-10的等级，得到

xxx\_module\_type属性，即模块等级。

* 对于模块内的属性，还可以通过聚类算法，找到属性间的关联，这里暂定名称为

xxx\_module\_inner\_type（模块属性内部关联），而inner\_type分出的类个数，将有参数调优决定。

* 合成的新表，名称为xxx\_module。

### 3.1风险统计

表3-1所示是风险模块各个建模及等级范围、数据类型、要求以及权重。

表3-1：risk\_module

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 | 权重 |
| risk\_module | is\_punish\_type | int | 默认值0 | 行政处罚等级1-5，缺失为0，下同 | 1 |
| is\_bra\_type | 经营异常等级1-5 | 1 |
| pledgenum\_type | 股权出质等级1-5 | 1 |
| is\_brap\_type | 行政处罚记录次数等级1-5 | 1 |
| tax\_unpaidnum\_type | 欠税等级1-5 | 1.5 |
| unpaid\_sum\_type | 欠缴保险额1-5 | 0.7 |
| is\_except\_type | 异常等级1-5 | 1 |
| declaredate\_type | 最新纠纷日期1-5 | 0.2 |
| appellant  \_amount\_type | 原告总数等级 | 0.3 |
| defendant  \_amount\_type | 被告总数等级 | 0.6 |
| enforce\_amount  \_type | 执行金额等级 | 1.8 |
| record\_date  \_type | 执行日期等级（越大等级越高） | 0.3 |
| judge\_new\_count  \_type | 诉讼次数等级 | 1 |
| is\_justice  \_credit\_type | 工商部失信等级 | 1 |
| is\_justice  \_creditaic\_type | 司法风险失信等级 | 0.6 |
| risk\_module | double | 默认值0.0 | 经营风险模块加权总分 | / |
| risk\_module\_type | int | 默认值0 | 风险等级1-10 | / |
| risk\_module  \_inner\_type | 风险模块内部聚类 | / |

### 3.2投资统计

表3-2所示是投资模块各个指标等级范围、数据类型、要求以及权重。

表3-2：investment\_module

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 | 权重 |
| investment  \_module | insurance\_num\_type | int | 默认值0 | 年平均参保等级 | 0.3 |
| bidnum\_type | 中标等级 | 2 |
| branchnum\_type | 企业分支等级 | 2.5 |
| subconam\_total\_type | 认缴总额等级 | 1 |
| liacconam\_type | 累计实缴等级 | 0.5 |
| lisubconam\_type | 累计认缴等级 | 0.5 |
| investnum\_type | 投资次数等级 | 2.5 |
| shopnum \_type | 网店个数等级 | 3 |
| investment\_module | double | 默认值0.0 | 投资加权总分 | / |
| investment\_module \_type | int | 默认值0 | 投资等级，等级1-10 | / |
| investment\_module  \_inner\_type | 投资模块内部聚类 | / |

### 3.3知识产权统计

表3-3所示是知识产权模块各个指标等级范围、数据类型、要求以及权重。

表3-3：creativity\_module

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 | 权重 |
| creativity\_module | ibrand\_num\_type | int | 默认值0 | 商标申请次数等级 | 1 |
| icopy\_num\_type | 软著登记等级 | 1 |
| ipat\_num\_type | 专利申请等级 | 1 |
| idom\_type | 域名知识产权等级 | 1 |
| creativity\_module | double | 默认值0.0 | 知识产权(创新模块)总分 | / |
| creativity\_module\_type | int | 默认值0 | 创新等级 | / |
| creativity\_module  \_inner\_type | 知识产权（创新模块）内部聚类 | / |

### 3.4品牌统计

表3-4所示是品牌模块各个指标等级范围、数据类型、要求以及权重。

表3-4：brand\_module

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 | 权重 |
| brand\_module | is\_jnsn\_type | int | 默认值0 | 济南市中专精小企业等级 | 1.3 |
| level\_rank\_type | 科技等级 | 1.5 |
| is\_infoa\_type | 驰名商标 | 1.5 |
| is\_infob\_type | 著名商标 | 1 |
| passpercent\_type | 质检通过率 | 0.7 |
| brand\_module | double | 默认值0.0 | 品牌模块总分 | / |
| brand\_module  \_type | int | 默认值0 | 品牌等级1-10 | / |
| brand\_module  \_inner\_type | 品牌模块内部聚类 | / |

### 3.5招聘统计

表3-5所示是招聘模块各个指标等级范围、数据类型、要求以及权重。

因为三张表已经汇总过了，所以不用再额外操作，只需要直接对三大平台的招聘数做聚类打分即可，由于只对招聘总数做评估，所以不做内部属性聚类。

表3-5：recruit\_module

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 | 权重 |
| recruit\_module | qcwynum | int | 默认值0 | 前程无忧招聘 | 1 |
| zhycnum | 中华英才招聘 | 1 |
| zlzpnum | 智联招聘 | 1 |
| recruit\_module | 三个平台招聘总数 | / |
| recruit\_module\_type | 招聘数等级1-10 | / |

### 3.6信用统计

表3-6所示是信用模块各个指标等级范围、数据类型、要求以及权重。

由于守合同重信用企业的数据量相比信用等级数据量少了太多，且本身数据属性维度仅为2维，只对信用模块加权总分做聚类，不进行内部聚类。

表3-6：credit\_module

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 | 权重 |
| credit\_module | is\_kcont\_type | int | 默认值0 | 守合同重信用企业等级 | 1.3 |
| credit\_grade  \_type | 信用等级 | 1 |
| credit\_module | double | 默认值0.0 | 信用模块加权总分1-10 | / |
| credit\_module\_type | int | 默认值0 | 信用等级 | / |

### 3.7企业基本信息统计

表3-7所示是基本信息模块各个指标等级范围、数据类型、要求以及权重。

由于注册资本权重很高，内部分类实际就是企业基本信息等级分类，所以不做内部分类。

表3-7：company\_baseinfo\_module

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 | 权重 |
| company\_baseinfo  \_module | regcap\_type | int | 默认值0 | 注册资本等级 | / |
| company\_baseinfo  \_module | double | 默认值0.0 | 基本信息模块总评 | / |
| company\_baseinfo  \_module\_type | int | 默认值0 | 企业基本信息等级 | / |

## 模块总评

经过了步骤三的模块统计后，企业的数据已经成功被压缩到了7个维度，接下来需要将各个module\_type通过加权求和的方式，获得企业的总评分，并再使用K-Means的方式获得其总评等级（此处分为10个等级），并再根据7个维度的module\_type，直接做K-Means训练，挖掘每个属性之间的内部关联，获得ent\_inner\_type，得到最终表ent如表4-1所示：

表4-1：ent

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 表名 | 属性名 | 类型 | 特殊要求 | 备注 | 权重 |
| ent | risk\_module\_mark | int | 默认值0 | 1-风险 | -1.11 |
| investment\_module\_mark | 2-投资 | 1 |
| creatity\_module\_mark | 3-创新能力 | 1 |
| brand\_module\_mark | 4-品牌 | 1 |
| recruit\_module\_mark | 5-招聘 | 1 |
| credit\_module\_mark | 6-信用 | 1 |
| base\_info\_module | 7-基本信息 | 1 |
| ent | double | 默认值0.0 | 企业总分 | / |
| ent\_type | int | 默认值0 | 企业总分等级，10个等级 | / |
| ent\_inner\_type | 分成若干个类 | / |