**云啸智推一期**

**基础算法设计说明**

目录

[1 数据演算模块 3](#_Toc529205768)

[1.1 模型规则标签演算 3](#_Toc529205769)

[1.1.1 积分引擎 3](#_Toc529205770)

[1.1.2 积分模型 3](#_Toc529205771)

[1.1.3 积分模型配置 3](#_Toc529205772)

[1.1.4 积分计算 3](#_Toc529205773)

[1.2 数据云池统计演算 4](#_Toc529205774)

[1.2.1 数据统计配置 4](#_Toc529205775)

[1.2.2 数据统计计算 4](#_Toc529205776)

[1.2.3 数据监测 4](#_Toc529205777)

[1.3 热词演算 4](#_Toc529205778)

[1.3.1 热词统计配置 4](#_Toc529205779)

[1.3.2 热词统计计算 4](#_Toc529205780)

[1.3.3 热词订阅配置 5](#_Toc529205781)

[1.3.4 热词订阅计算 5](#_Toc529205782)

# 数据演算模块

数据演算模块：

——模型规则标签演算：基于模型规则，数据抽取演算打标

——数据云池统计演算：所有数据综合汇总，根据时间、区域、分类进行演算统计

——热词演算：对数据进行分词演算统计，根据词频推送出关键热词

## 模型规则标签演算

积分引擎通过加载积分模型配置，对各积分模型进行积分计算。积分模型通过平台管理系统的滤网管理进行配置管理，其滤网名称就是标签。

### 积分引擎

积分引擎负责处理管理对象的滤网积分，每一个分类积分的分值由多个积分模型共同计算得到，反映了管理对象在社会生活中某一块的积分指数。积分模型在计算分类积分时的所占的比重可以通过平台管理系统方便地进行调整，以适应外部环境的变化。

### 积分模型

积分模型由一组在大数据分析框架下开发实现的算法模型组成，每一个积分模型都用于分别计算管理对象在某一特定维度下的滤网行为（规则）上的积分指数。

### 积分模型配置

积分模型配置用于配置每一个积分模型规则的可选参数。配置信息使用MySQL数据库保存，后期有新的积分模型加入或原有模型需要针对新的环境做调整时只需要增加或修改配置记录即可。

每一个积分模型规则都有自己的模型配置信息，配置信息记录了该模型在计算时的一些可选参数及满足该模型的管理对象应该获得的单项积分权重。

### 积分计算

积分引擎在进行积分计算时首先调用所有积分规则进行单项积分计算，然后对每一个管理对象，将其获得的所有的单项积分按权重累加后转换为百分制的分类总积分。

## 数据云池统计演算

统计引擎加载数据统计配置，对每个统计任务做数据统计计算；监测引擎监控统计任务的执行。

### 数据统计配置

统计配置用于配置统计任务的具体内容。配置信息使用MySQL数据库保存。

统计任务按数据分类，数据类别，数据种类的层级划分；执行统计任务时通过定义具体数据源的辖区分类字段（派出所、网格、地区、时间）做分组统计；过滤基础数据可通过配置具体数据源的条件。

### 数据统计计算

统计引擎在进行每个统计任务的统计计算时，数据按配置条件过滤，按辖区分类字段分组统计，再将统计任务结果进行保存。

### 数据监测

监测引擎读取数据统计配置，通过查看任务结果来检测统计任务是否执行。

## 热词演算

统计引擎加载热词统计配置，对每个热词任务做分词统计计算；订阅引擎加载订阅配置，根据规则推送信息。

### 热词统计配置

统计配置用于配置统计任务的具体内容。配置信息使用MySQL数据库保存。

统计任务提供数据源及其定义的分词统计字段，过滤基础数据可通过配置具体数据源的条件。

### 热词统计计算

统计引擎对每个热词任务做分词统计计算时，数据按配置条件过滤，排除冷词库的词，再将统计结果按词分组累加进行保存。

### 热词订阅配置

订阅配置用于配置订阅任务的具体内容。配置信息使用MySQL数据库保存。

### 热词订阅计算

订阅引擎根据订阅规则（多个热词的交并集关系），推送到制定的用户消息列表，如果有配置配置方式，还可推送到指定短信或邮箱。