

ICS 35.240.99

L 67

# ZWFW

## 国家政务服务平台标准

C 0127-2018

---

### 国家政务服务平台 政务服务大数据分析模型指南

2018-12-28 发布

2018-12-28 实施

---

国务院办公厅电子政务办公室 发布



# 目 次

前言 .....	III
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 建模流程.....	2
4.1 概述.....	2
4.2 目标制定.....	2
4.3 数据预处理.....	2
4.4 模型构建与管理.....	2
4.5 结果输出.....	3
5 数据分析方法.....	3
5.1 概述.....	3
5.2 基础分析方法.....	3
5.3 数据挖掘方法.....	3
5.4 空间分析方法.....	4
6 分析指标.....	4
6.1 概述.....	4
6.2 基本原则.....	4
6.3 指标构成.....	4
6.3.1 服务事项.....	4
6.3.2 办件情况.....	5
6.3.3 数据交换.....	5
6.3.4 用户分析.....	5



## 前 言

本标准按照 GB/T 1.1-2009 给出的规则起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本标准由国务院办公厅电子政务办公室提出并归口。

本标准起草单位：国务院办公厅电子政务办公室、浙江省人民政府办公厅、广东省人民政府办公厅、江西省人民政府办公厅、四川省人民政府办公厅、南京市政务服务管理办公室、中国电子技术标准化研究院。

本标准主要起草人：卢向东、尹智刚、陈治佳、马运领、佟舟、姚远、孔令军、徐云、李景曦、王赟萃、李松渊、孙杨、张军、钱学文、李恒训、赵菁华、陈亚军。



# 国家政务服务平台政务服务大数据分析模型指南

## 1 范围

本标准规定了政务服务大数据的建模流程、分析方法和分析指标。

本标准适用于指导政务服务工作人员依托国家政务服务平台进行数据分析建模，同时引导各地区各部门开展政务服务大数据分析工作。

## 2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 25647-2010 电子政务术语

GB/T 35295-2017 信息技术 大数据 术语

C 0130-2018 国家政务服务平台政务服务数据整合要求

## 3 术语和定义

GB/T 25647-2010 和 GB/T 35295-2017 界定的以及下列术语和定义适用于本文件。

### 3.1

**分析模型 analysis model**

对客观事物或现象的一种描述。通过分析其相互作用机制，揭示内部规律，根据理论推导、对观测数据的分析和实践经验等，设计一种模型来代表所研究的对象。

### 3.2

**变量 variable**

描述观察单位或个体的各项特征的量。

### 3.3

**数据挖掘 data mining**

从大量的数据中通过算法搜索隐藏于其中信息的过程。

### 3.4

**指标 indicator**

客观的反映现象总体或样本数量特征的范畴及其具体数值。

### 3.5

**特征 feature**

反映被观察的对象或现象在某方面的性质或特性。

## 4 建模流程

### 4.1 概述

依托政务服务大数据分析工具，根据政务服务业务需求构建大数据分析模型。建模流程包括目标制定、数据预处理、模型构建与管理 and 结果输出。当建模过程包含多个算法时，循环执行数据预处理和模型构建步骤。建模流程见图 1。

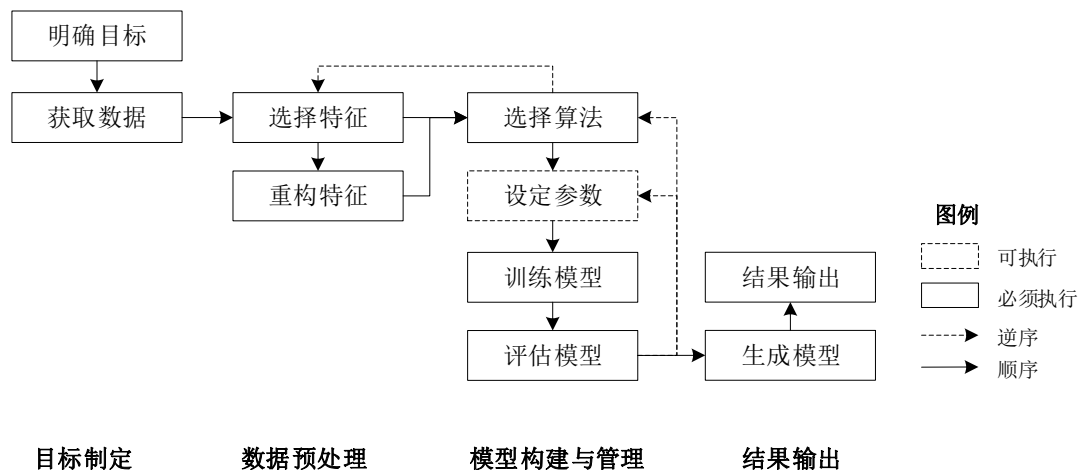


图 1 建模流程

### 4.2 目标制定

根据政务服务大数据分析需求制定建模目标，获取分析数据。分析数据内容包括但不限于事项信息、办件信息、证照目录信息、数据交换日志信息、用户及用户行为信息和评价信息。数据可来源于 C0130-2018 中第 7 章、第 8 章整合后的政务服务数据立方体或政务空间数据。

### 4.3 数据预处理

根据政务服务建模目标确定分析特征。当特征不能直接用于建模时，需对特征进行重新选择或重构。特征重构方法包括数学函数变换、归一化、连续变量离散化、特征组合等特征变换操作。

### 4.4 模型构建与管理

利用特征和算法构建政务服务数据分析模型，并通过调整参数优化模型。实现分析模型的统一管理，包括模型部署、业务验证、模型监控和模型定制等。模型构建过程包括选择算法、设定参数、训练模型和评估模型。

- a) 选择算法。根据数据分析需求选择相关分析方法，分析方法包括聚类分析、分类分析、时间序列分析、回归分析等。数据分析方法见第 5 章；
- b) 设定参数。若分析方法不需要设定参数，可直接加载算法输出结果，如线性回归分析。若分析方法需设定参数，应按相关方法给定初始化参数，如聚类分析中的 K 均值算法，需给出类别数量、起始聚类中心和迭代次数上限等参数；
- c) 训练模型。利用选定的特征、算法和参数执行运算，对模型进行训练；
- d) 评估模型。根据输出结果和相关评价方法，判断结果是否满足需求。若不能满足需求，应调整参数、重构特征或重新选择算法。

## 4.5 结果输出

当结果能满足业务人员的分析需求时，记录输入特征、方法、初始化参数等内容，保存模型。

## 5 数据分析方法

### 5.1 概述

政务服务大数据分析为业务人员面向政务服务业务开展数据统计分析、多维分析、深度挖掘和空间分析提供方法支持。数据分析方法包括基础分析方法、数据挖掘方法和空间分析方法。

### 5.2 基础分析方法

基础分析方法用于政务服务统计分析和多维分析，包括但不限于：

- a) 全表统计。对政务服务数据进行单一变量或组合变量的描述性统计分析，包括总和、平均值、最大值、中位数、最小值、方差、标准差、极差等。如计算各省份办件量总数、均值、中位数，可横向分析全国办件情况；
- b) 比率。两数相比所得的值，如本月办件量同比，即今年本月办件量与去年本月办件量的比率；
- c) 排序。将杂乱无章的数据，通过一定的方法按某种规律进行排列。如法定办结时限最长事项，即将事项按法定办结时限降序排列获取。

### 5.3 数据挖掘方法

数据挖掘方法用于开展深度政务服务业务分析工作，包括但不限于：

- a) 分类分析。将一组数据对象按不同的分类模式划分为不同的类别。分类分析算法包括朴素贝叶斯分类、决策树分类、逻辑回归分类、随机森林分类、支持向量机分类、梯度渐进树分类等。如根据用户和用户行为信息对用户分类，可提供精细化服务；
- b) 聚类分析。按相似性和差异性将一组对象划分成若干类，使类内相似度较高，类间相似度较低。聚类分析算法包括K均值聚类、基于密度的聚类等。如对用户访问政务服务页面的栏目聚类，可定性分析用户关注内容；
- c) 关联分析。描述数据项之间所存在关系的规则，挖掘隐藏在数据间的关联或相互关系。关联分析算法包括Apriori算法、FP-Growth算法等。如对用户年龄和办理事项主题进行关联分析，了解各年龄段用户热门办理事项；
- d) 回归分析。通过建立回归方程，产生一个将数据项映射到一个预测数据项的函数，发现变量或属性间的依赖关系。回归分析算法包括线性回归、广义线性回归、决策树回归、随机森林回归、生存回归、保序回归等。如通过构建某地区企业注册量与经济、人口、区位、政策等因素的回归方程，探查该地区企业注册的影响因素；
- e) 时间序列分析。根据客观事物发展的连续规律性，运用历史数据推测未来的发展趋势，对具有时间顺序关系的数据进行分析。时间序列分析算法包括指数平滑法、自回归滑动平均模型、隐马尔可夫模型、条件随机场等。如对办件量、用户量进行分析，预测未来办件量、用户量的变化趋势，或利用隐马尔可夫模型、条件随机场抽取工单中的人名、机构名、地名等实体信息，进行序列模式分析。

## 5.4 空间分析方法

空间分析方法用于开展与空间位置相关的政务服务业务分析，包括但不限于：

- a) 空间查询。利用空间索引机制，找出符合特定条件的空间数据。空间查询方法包括几何查询、属性查询等。如查询某地区政务服务办事大厅位置；
- b) 缓冲区分析。以点、线、面要素为基础，建立其周围一定范围内的缓冲区多边形图层。如分析用户 5 公里范围内的政务服务大厅；
- c) 叠加分析。在统一空间参考系统下，将同一地区两个地理对象的图层进行叠置，以产生空间区域的多重属性特征，或建立地理对象间的空间对应关系。叠加分析支持逻辑运算、数学运算、函数运算等几何运算方法，支持点与多边形叠加、线与多边形叠加、多边形叠加、栅格图层叠加等叠加方法。如将县级行政区划图和政务服务用户分布图进行叠加，可分析用户在各县的分布情况；
- d) 分区统计分析。对不同统计区的相同统计指标，用颜色、图形或图表的形式展示统计指标的数量和组成。分区统计分析方法包括定位统计图法、分区统计图法、分级统计图法等。定位统计图法以图形或图表的形式描绘统计指标的数量和组成；分区统计图法、分级统计图法用颜色表示统计指标的数量差别。如在地图上以柱状图的形式展示各省份的办件量；
- e) 核密度分析。用于计算要素在其周围邻域中的密度，以特殊高亮的形式显示热点地理区域。如用户空间分布的热点区域或法人单位空间分布的热点区域；
- f) 热点分析。统计识别具有统计显著性的热点和冷点。如办件较为集中的热点区域或用户较为集中的位置。

## 6 分析指标

### 6.1 概述

分析指标是实施数据分析、建立数据分析模型的基础，通过设计和构建“涵盖面广、层次分明”的政务服务数据统计分析指标，描述和评价政务服务能力，及时掌握政务服务平台运行情况。分析指标包括但不限于服务事项、办件情况、数据交换和用户分析。

### 6.2 基本原则

分析指标的构建原则包括但不限于：

- a) 实用性。分析指标来源于基础性的统计工作，应满足业务人员的基础性和综合性分析需求；
- b) 规范性。分析指标的计量单位、统计范围、统计频度等要素定义符合规范性要求，分析指标及相关要素设置定义清晰；
- c) 科学性。确保指标信息来源符合规范，设计的数据、文字等信息客观严谨，经得起实践检验；
- d) 可扩展性。在确保基础架构相对稳定的基础上，当数据更新和分析需求变动时，分析指标具有一定的扩展性。

### 6.3 指标构成

#### 6.3.1 服务事项

政务服务事项客观展示政府为企业、群众办事提供服务的能力。

服务事项分析指标包括全国基本目录事项、全国实施清单事项数量、面向自然人事项数

量、面向法人事项数量、各地区实施清单事项数量、各部门实施清单事项数量、各类型事项数量、各层级事项数量、各实施主体事项数量、不同材料种类事项数量、全国网上办理率、各地区网上办理率、各部门网上办理率、承诺累计节省的天数、承诺平均节省的天数、事项办理环节平均时长、事项办理环节统计、收取材料数量和到办事现场次数等。

### 6.3.2 办件情况

办件情况揭示企业、群众与政府服务的互动规律。

办件情况分析指标包括办件申请量、办件受理量、办件办结量、即办件数量、承诺件数量、自然人办件量、法人办件量、各行使层级办件量、各事项类型办件量、各类实施主体办件量、各类服务对象办件量、各年龄段办件量、超出法定时限的办件量、超出承诺时限的办件量、全国办结率、各地区办结率、各部门办结率、实际办件累计节省的时间、实际办件平均节省的时间、平均单环节耗时、高频办理事项、低频办理事项、自然人热门办理主题、法人热门办理主题和各年龄段热门办理主题等。

### 6.3.3 数据交换

数据交换揭示政务服务平台间数据交互情况，反映全国一体化在线政务服务平台内部运行机制。

数据交换分析指标包括数据总量、事项数据总量、办件数据总量、证照数据总量、用户数据总量、各地区各部门平台接入情况、各地区各部门数据交换情况、数据服务接口调用量、证照请求量和证照下发量等。

### 6.3.4 用户分析

用户分析分析政务服务平台用户以及用户的访问使用行为，为平台改进提供参考。

用户分析分析指标包括访问量、浏览量、用户注册量、自然人用户量、法人用户量、实名制用户量、非实名制用户量、各年龄段用户量、各性别用户量、各时间段用户认证量、计算机端用户使用量、移动端用户使用量、活跃用户量、访问检索排行、热门访问时段、平均页面访问时间、平均跳出率、平均访问深度、访问去向和用户访问页面主题等。