

# 分层分类实现与应用

# 目录

## CONTENTS

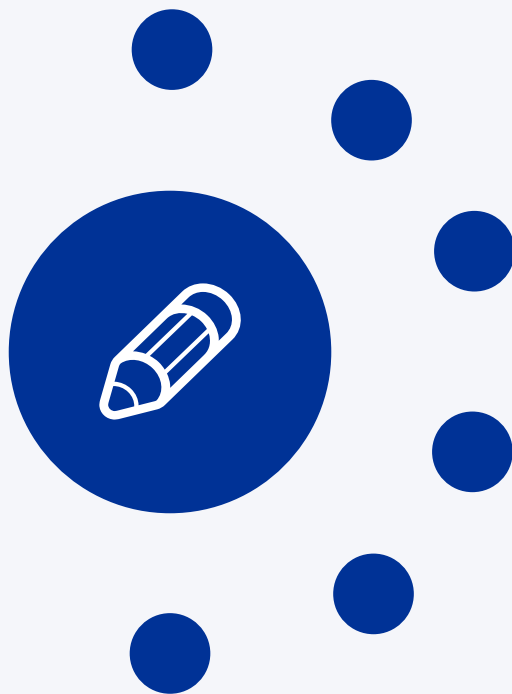
01 分层分类背景与介绍

02 分层分类算法实现

03 分层分类的应用



# 01 分层分类背景与介绍



- Linux是一种广泛使用的开源操作系统，由于不同的发行版面向不同的场景，具有不同的特性，包的开发和管理过程并不完全统一，这对Linux操作系统发行版的构建和维护提出了很大的挑战。
- 我们希望建立一套方法论，本方法论将遵循软件行业的基本原则，结合机器学习算法，基于 Linux 操作系统的元数据，开发出自动化的 Linux 架构分析工具。

分层分类模型在架构上将所有软件划分为五个层级

- 内核层 :在此层级中存放的软件包主要为内核及其直接依赖的软件包。
- 核心层 :该层定义了可扩展的最小人机交互环境。
- 系统层 :该层为操作系统提供了主要管理、扩展功能。
- 应用层 :该层定义了操作系统对用户以应用程序提供的功能。
- 使用场景层 : 该层可根据用户的不同需求划分出不同的使用场景。



形态分类

形态分类是根据软件包中所提供功能文件的形态不同，将软件包进一步细分为库、服务和工具三类。

功能分类

功能分类基于包所提供的功能进行的分类，每个软件包可以提供一个或多个功能信息，可以分为多个类别。

场景分类

软件包按照使用场景分类，根据用户使用该软件的常用场景进行分类。

形态元数据	描述
库	软件包所提供的功能以动态链接库或静态链接库形态呈现，被其他程序通过函数调用的方式来使用
服务	软件包所提供的功能以独立运行的后台进程形态呈现，其它进程通过进程间通信的方式来使用
工具	软件包的功能是以命令行界面或图形界面形态呈现，供用户交互使用

功能元数据标签	描述
通用	提供程序运行支持库功能为非管理用途的应用程序
网络	网络管理功能（如配置IP等）与对外提供网络服务的功能（如httpd）
系统管理	对账户和操作系统进行管理、配置、监控的功能
设备管控	用于管理打印机，扫描仪，键盘，鼠标，蓝牙等外设的软件包
视窗及渲染	包括2D、3D图形加速库库，如mesa等各个发行版所发行的桌面套件，如DDE，Xfce等
存储	磁盘管理、文件系统管理、远程存储，数据库等
容器和虚拟化	包含容器和虚拟化功能，如Docker、libvirt等
多媒体	图像、视音频数据的展示、播放、编辑、相关编解码库等功能
数据格式转换	通用格式的数据压缩，数据加解密等功能
其他	语言，纯文档类以及文档类型定义（DTD）等包

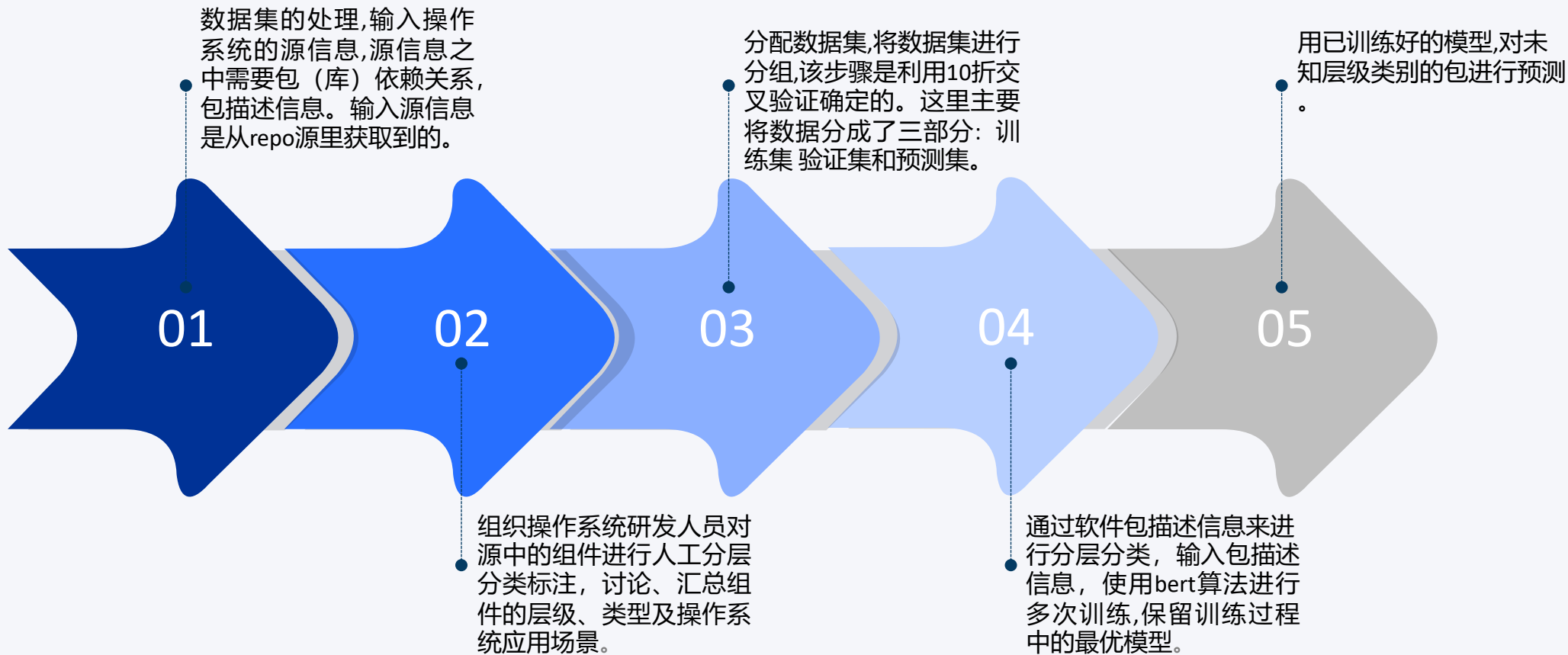
场景分类名称	主要软件包
数据库	MySQL、MariaDB、PostgreSQL
云原生	Docker、podman
大数据	Spark，Hadoop
人工智能	PyTorch、TensorFlow



## 02 分层分类算法实现

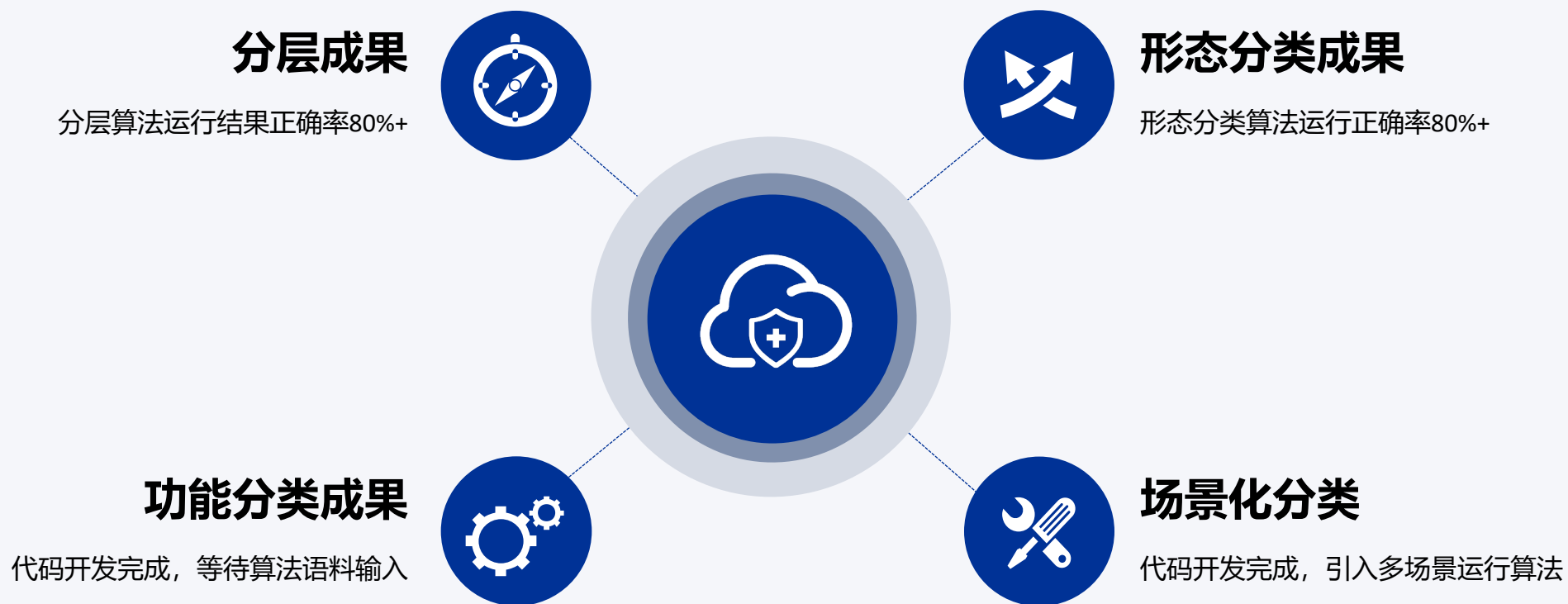


## 分层分类算法中主要利用bert进行组件的分层及分类



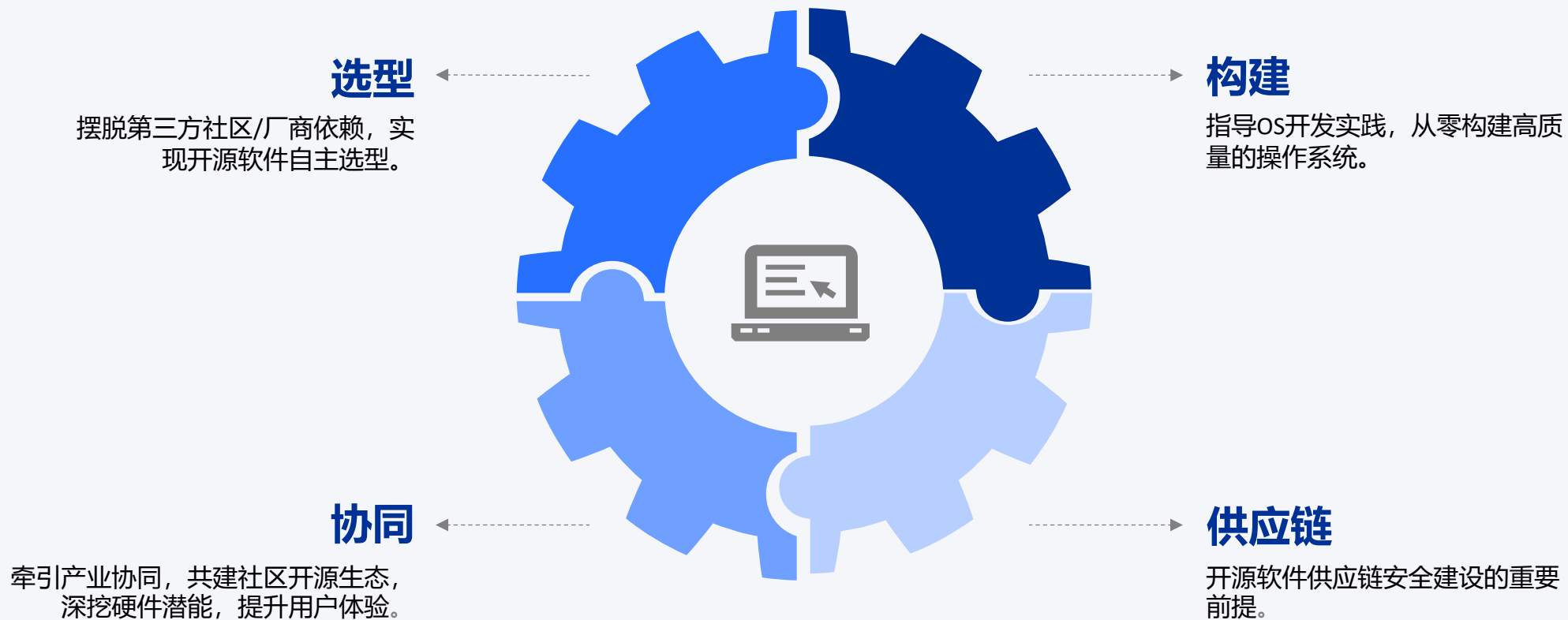


目前，我们已经对uos的6759 个二进制包、4749 个源码包进行了分层分类





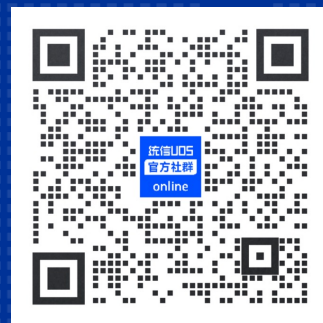
## 03 分层分类的应用



# 中国操作系统领创者 给世界更好的选择



统信软件官方微信公众号

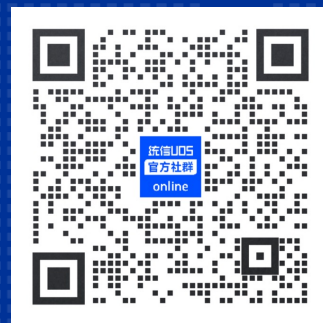


统信软件官方社群

# 中国操作系统领创者 给世界更好的选择



统信软件官方微信公众号

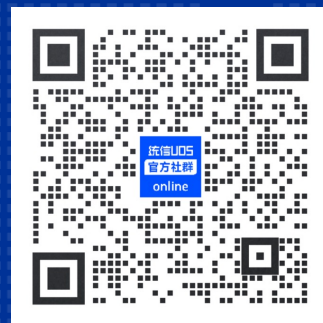


统信软件官方社群

# 中国操作系统领创者 给世界更好的选择



统信软件官方微信公众号



统信软件官方社群