

AI技术和云技术在视频监控中的应用

流式计算模型

陈一帅

yschen@bjtu.edu.cn

北京交通大学电子信息工程学院

内容

- 定义
- 原理
- 系统

流式计算模型

- 实时或近实时分析变得越来越关键
 - 视频流
 - 物联网
- 分析来自无限制流的数据的活动，即数据流分析
- 可追溯到 1990 年代在斯坦福，加州理工学院和剑桥等地进行的复杂事件处理的基础研究

内容

- 定义
- 原理
- 系统

节点预分析

- 数据流的分析有时需要靠近源
- 正在出现执行预分析的工具，其目的是确定应发送到云以进行更深入分析的数据子集
- 如 Apache Edgent 边缘分析工具，能够在 Raspberry Pi 等小型系统中运行

时间窗口

- 固定时间窗口：将输入流分为逻辑段，每段对应于一个指定的处理时间间隔。间隔不重叠
- 滑动窗口：允许窗口重叠。例如，窗口大小为 10 秒，每 5 秒启动一次
- 以会话为单位的窗口：将流划分为与数据的某些键相关的活动的会话（Session）。例如，某位用户的一连串鼠标点击可以捆绑到一起，作为一个时间上临近的系列点击会话
- 全局窗口：封装整个有界流

触发

- 与窗口相关联
- 触发对窗口内容的分析，并发布结果

内容

- 定义
- 原理
- 系统

系统

- Spark Streaming
- Apache Storm, 来自 Twitter 的 Heron
- Apache Flink, 来自德国 Stratosphere 项目
- Apache Beam, 来自 Google 的 Cloud Dataflow, 可以在 Flink, Spark, Google 云上运行
- Amazon Kinesis
- Azure Event Hubs
- IBM Stream Analytics

系统

- 亚马逊的事件流软件堆栈 Kinesis，包括三个服务
 - Kinesis Streams：提供有序的、可重播的实时流数据
 - Kinesis Firehose：支持极高规模的事件处理，可以将数据直接加载到 S3 或其他 Amazon 服务中
 - Kinesis Analytics：提供了基于 SQL 的分析工具，实时分析 Kinesis Streams 或 Firehose 中的流数据

小结

- 定义
- 原理
- 系统