

## 赛道 A 移动通信基站流量预测

随着移动通信技术的发展，4G、5G 给人们带来了极大便利。移动互联网的飞速发展，使得移动流量呈现爆炸式增长，从而基站的流量负荷问题变得越来越重要。一方面，在流量高峰期，大量基站呈现出负荷超过容量的问题，使得即使信号条件很好，网络速度也非常的慢，给用户带来非常差的体验。为了改善这个问题，需要给基站增加载频的数量来扩容，使基站可以承载更多的流量；另一方面，由于基站潮汐现象，使得在某些时段，用户数量会大幅降低。如图 1 所示，是某基站下行流量随时间变化的曲线。

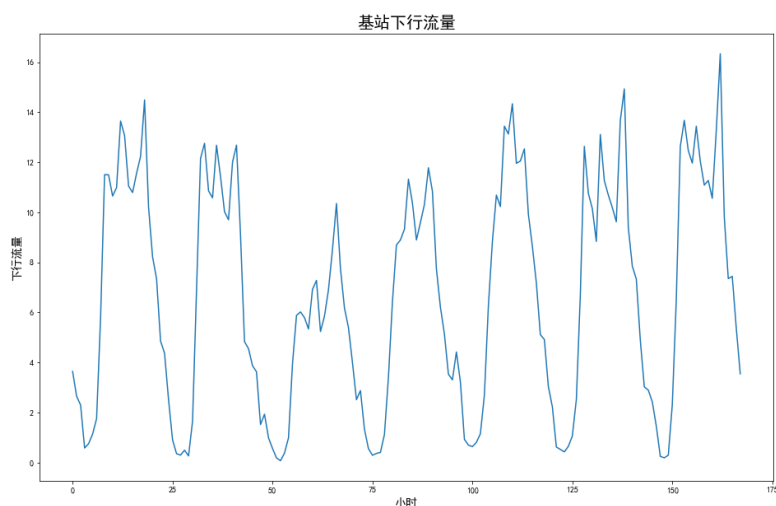


图 1 某基站下行流量随小时变化的曲线

在基站低流量时段，如果仍然按照高容量时段的载频数量来运行基站配置，会极大的浪费资源和能量。特别是现在每个城市的基站数量巨大，而且随着 5G 的不断部署，基站数量还会大幅增加。同时，每个基站的流量高峰和低谷的时段各不相同。如果所有基站都按照高容量时段来配置运行载频，则网络的能量消耗是非常巨大的。因此，需要根据流量的变化，计算

需要的载频数量，从而可以在不同时段打开或者关闭部分载频使得基站既可以满足对用户的服务，又可以尽可能低的消耗能量和资源。

由于基站数量巨大，无法通过人工实时关注每个基站的流量变化，需要给每个基站设置根据时段自动开关载频的程序。这样就需要知道一段时间内基站流量关于时段的变化值，特别是基站在每个小时的上下行流量值。从而可以知道基站在每个时段需要的载频的数量，进而设置一定时间内基站载频自动开关的程序。

另外，无论基站流量随时段怎么波动，从长期来看，大部分基站的整体流量是呈逐渐增加趋势的。当整体流量增加到一定程度时，这种动态开关载频的方式已经无法满足基站在流量峰值时候的需求了，因此需要做物理扩容，新建扇区或者新建基站。由于物理扩容需要涉及采购、费用、总体布局等问题，因此规划需要的时间非常长，所以需要提早预估出基站需要物理扩容的时间，从而可以更早的进行规划和设计。

基于以上背景，请你们的团队根据附件给出的数据，通过数据分析与建模的方法帮助移动通信运营商解决下面的问题：

**问题 1：**对小区(基站的每个扇区)的上行和下行流量随时间的变化进行建模，并用附件 1 中的“训练数据”(部分小区 2018 年 3 月 1 日至 4 月 19 日的小时级流量数据)训练模型，给出各个小区小时级上行和下行流量的预测模型，预测这些小区后面一周(4 月 20 日至 4 月 26 日)的小时级流量变化，并将附件 2“短期验证选择的小区数据集”中涉及的小区数据填充完成，填充结果提交到竞赛网站平台，请勿改变附件 2 中数据的格式及小区 ID 编号顺序。

**问题 2:** 预测小区上下行流量的长期变化趋势，给出每个小区长期流量预测模型，并对附件 3 “长期验证选择的小区数据集”中涉及的小区，预测其在 2020 年 11 月 1 日至 11 月 25 日每天的总的上行和下行流量，填充该数据表，并提交到竞赛网站平台，请勿改变附件 3 中数据的格式及小区 ID 编号顺序。

**问题 3:** 依据给出的样本数据集，你们觉得还有哪些问题值得研究，并给出你们的思路？

附件 1: 训练数据.csv;

附件 2: 短期验证选择的小区数据集.csv;

附件 3: 长期验证选择的小区数据集.csv。