队伍编号	1771
赛道	（A）

大规模移动通信基站时序多特征分类和休眠方法研究

摘 要

近年来，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，基站建设的规模越来越庞大，同时基站的流量负荷问题变得越来越重要。因此，我们该如何准确对大规模基站进行分类是下一步针对性和分批性进行管理的重要步骤之一，以及给每个基站设置根据时段自动开关载频的程序以及相应的休眠策略是有效优化基站运行的重要手段之一。本研究中总共给了 13 万+基站的数据，问题 1 主要针对这些基站进行分类并阐述相应类别的特点，问题 2 针对优化基站的运行设置合适的阈值和制定优越的策略。

针对问题一，主要实现基站的分类和相应类别特点的阐述。首先在数据层最需要进行的是数据得预处理和特征数据处理，前者主要包括基站提取、时间戳处理、数据排序及清洗和时间段划分及统计，最终筛选出 **118756 条可用数据**；后者提取上下行流量高峰时段（Top3）、日均流量（GB）、日标准差、日峰值（GB）、连续变化绝对值之和以及上下行流量均值比作为小区基站历史流量数据的**时序特征**，共计 **30 维特征**。归一化后，输入到模型层，利用**改进后的模糊 C 均值（FCM）**非监督分类方法进行分类，其改进点在于首先利用类间距离和类内距离确定**最佳分类数目为 12**。而后，通过分类结果，计算各类别基站下数据的**众数或者中位数**用以 12 类基站特点的分析。

针对问题二，主要针对优化基站的运行设置合适的阈值和制定优越的策略。本文以**“基站休眠阈值的设置—基于长短期记忆网络（LSTM）的短期预测模型—基站休眠策略的制定”**为研究主线，构建了基于流量负载预测的大规模移动通信基站休眠方法框架。其中，通过归一化负载限制值，将网络负载水平划分为四个等级，即**超低负载水平、低负载水平、中负载水平**和**高负载水平**。提前设置相应的休眠比例参数，根据短期预测结果，确定基站在未来各小时段内应采取休眠或者唤醒的状态，达到优化基站运行能耗和效率的目的。

本研究提出的方法可有效用以大规模基站非监督分类及优化基站运行的阈值设置和策略制定，为超大规模移动通信基站的布设和后续优化管理提供了有效的基础保障。

关键词：基站分类；时序多特征；模糊 C 均值（FCM）；类内及类间距离；LSTM 预测模型；基站休眠策略

目　录

一、引言

1.1 问题背景及相关工作

随着城市化现象越来越明显，手机电脑等智能设备的普及，用户对于通讯网络业务的需求逐渐增大。基站作为移动通讯、网络服务的载体，为上述设备提供有效的通信支撑。然而，频繁的数据传输在高密集度的活动区域中会对基站运营造成巨大的压力[1-2]。近年来，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[3]。

因此，在基站低流量时段，特别是在如今基站数量庞大和 5G 通讯逐步普及的时候，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。另外，通过对一天内各个基站上行流量和下行流量的观察，通常可以发此案每个基站的流量高峰和低谷的时段各不相同[4]。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。针对上述问题，引申出两个亟需解决的挑战：其一，如何对大规模基站进行分类管理；其二，如何有效优化基站运行时段，设置合适的基站休眠/唤醒状态。

对于基站分类问题，刘濛等人提出将小蜂窝基站按照“活跃度”进行分类，采用模糊分层聚类的方法对蜂窝基站进行相似度聚类。该方案可以优先在“活跃度”较高的蜂窝基站群中分配并复用物理小区识别（Physical Cell Identification，PCI），这样既考虑了用户的 QoS（Quality of Service，QoS），也提高了 PCI 的利用率[5]；，针对 PCI 数量有限以及冲突和混淆的问题，有很多新颖的解决方案被提出。文献[6]中利用“博弈论”优化了 PCI 分配的效率，尽管这种方案提高了 PCI 的复用效率，但是分配方案较为复杂，时间复杂度高，耗时严重，难以适用于热点地区大规模部署的小蜂窝环境。文献[7]采用夹角余弦值距离法对蜂窝小区基站进行模糊聚类，优先对活跃度高的小区基站分配 PCI，蜂窝基站的相似性体现在不同数据的方向属性上，忽略了数据的数值大小，缺乏对蜂窝

基站活跃度参数的综合考量。

对于如何优化基站运行效率，便产生了另外一个问题，我们该如何准确预测基站在某一时间段内上行流量和下行流量的数值，以给每个基站设置根据时段自动开关载频的程序。并且，从长期来看，上下行流量的数值是逐渐扩大的，我们该如何预先知道数额，以达到给基站做提前物理扩容的规划和设计[8]。文献[9]中提出了一种基于机器学习的流量预测和基站休眠方法，仿真结果显示，所提出的休眠方案，在保障用户服务质量，将中断概率保持在较低水平的同时，能够降低网络能源消耗、使网络能效具有明显提升。此外，还有许多基站休眠技术被提出，文献[10]提出了基于基站休眠的蜂窝网络能量效率优化技术，文献[11]提出了 5G 密集异构网络下的基站休眠技术研究方法。

针对现有基站分类方法难以自动确定分类数目的问题，本文拟引入多目标综合研判方法，将类内距离和类间距离[12]同时考虑，确定最佳的非监督分类数目。同时借鉴已有基站休眠策略制定的流程，以“基站休眠阈值的设置—基于长短期记忆网络（LSTM）的短期预测模型—基站休眠策略的制定”为研究主线，拟构建了基于流量负载预测的大规模移动通信基站休眠方法框架。

1.2 问题重述

问题 1：对附件 1（见初赛题目）的数据进行预处理，提取附件 1 关于上行和下行流量时间序列数据的特征，依据你们所提取的特征对附件 1 所给的小区进行分类（类别不超过 30 个），并说明每一类的特点。

问题 2：基站运行具有潮汐现象，如果按高容量时段的载频数量来运行基站配置会造成浪费。根据流量变化设置自动开关限制部分载频的方法，可以有效优化基站的运行，节约能源。基站开关载频的流量阈值则成为研究的关键，过低会造成用户使用的体验感差，过高仍然会造成资源的浪费。请你们给出阈值的设置策略和具体结果。

将问题 1、2 的解答写成一篇论文，明确你们的思路、模型、方法和结果。

二、问题分析与思路

根据上述问题背景和问题重述，我们需要通过数据挖掘以及数学建模方式帮助移动通信运营商更好地掌握不同区域的业务流量特征模式，为其设置基站及提出相应的业务服务提供数据支持。除此以外，我们需要针对每个基站的不同历史观察数据制定出合理的休眠策略，在满足用户流量需求的前提下减少能耗，降低运营成本。首先我们需要对历史业务量的观测数据进行数据处理，整理时间戳信息，清洗缺失冗余数据及样本量过

小的小区基站信息。在此基础上针对问题 1 及问题 2 开展进一步研究分析。

问题 1：主要需要解决 1）上下行流量时间序列特征提取及 2）小区基站分类两个问题。根据预处理后的 2018 年 3 月 1 日至 4 月 19 日的小时级历史流量数据，我们将其划分成工作日及非工作日两个部分，并将每个部分按小时进行统计，最终得到 48 个时间区间。此处，我们考虑将上下行流量的高峰时段（记录前三个高峰值）、日均流量、日流量标准差、日流量峰值、日流量连续变化绝对值之和和上下行流量之比作为特征指标。小区基站分类本质上是无监督聚类问题，本文中我们考虑采用自动确定聚类数目后改进的模糊 C 聚类（FCM）方法在时序特征基础上对小区基站进行聚类。根据聚类结果中小区基站的指标分布进一步对小区的流量模式进行标注说明。

问题 2：主要针对优化基站的运行设置合适的阈值和制定优越的策略。本文以“基站休眠阈值的设置—基于长短期记忆网络（LSTM）的短期预测模型—基站休眠策略的制定”为研究主线，构建了基于流量负载预测的大规模移动通信基站休眠方法框架。其中，通过归一化负载限制值 β_1，β_2，β_3，将网络负载水平划分为四个等级，及超低负载水平、低负载水平、中负载水平和高负载水平。提前设置相应的休眠比例参数（z_1、z_2、z_3、z_4），根据短期预测结果，确定基站在未来各小时段内应采取休眠或者唤醒的状态，达到优化基站运行能耗和效率的目的。

三、大规模移动通信基站时序多特征非监督分类和基于流量负载预测的基站休眠方法

表 1. 本文所用符号说明

符号	含义	符号	含义
U_t	t 时间段的业务流量	SSE	误差平方和
avg_{up}	上行日均业务流量	Sep	计算类间距离
avg_{down}	下行日均业务流量	N	样本点数目
x_j	第 j 个样本点	μ_{ij}	模糊隶属度矩阵
U_i	第 i 个聚类中心	U_k	第 k 个聚类中心
f_t	点乘法运算 1	$\tilde{C}_t$	新的候选解
o_t	点乘法运算 2	h_t	LSTM 输出
i_t	点乘法运算 3	LT	流量负载
β_1，β_2，β_3	归一化负载限制值	z_1、z_2、z_3、z_4	休眠比例参数

表 1 为本研究中所用到的符号及其含义，另外，在该研究中，有一个重要假设需要

说明，其为：假设各个基站之间的上行流量及下行流量是互相独立的，不存在依赖关系。针对大规模移动通信基站的分类问题，我们提出了一种基于时序多特征的非监督分类方法；针对如何有效优化基站的运行，节约能源，基站开关载频的流量阈值则成为研究的关键，过低会造成用户使用的体验感差，过高仍然会造成资源的浪费。我们设计了基于流量负载预测的大规模移动通信基站休眠方法用以优化基站的运行，主要包括：短期预测模型、基站休眠阈值设置和休眠策略制定。上述具体方法模型的框架如图 1 所示。

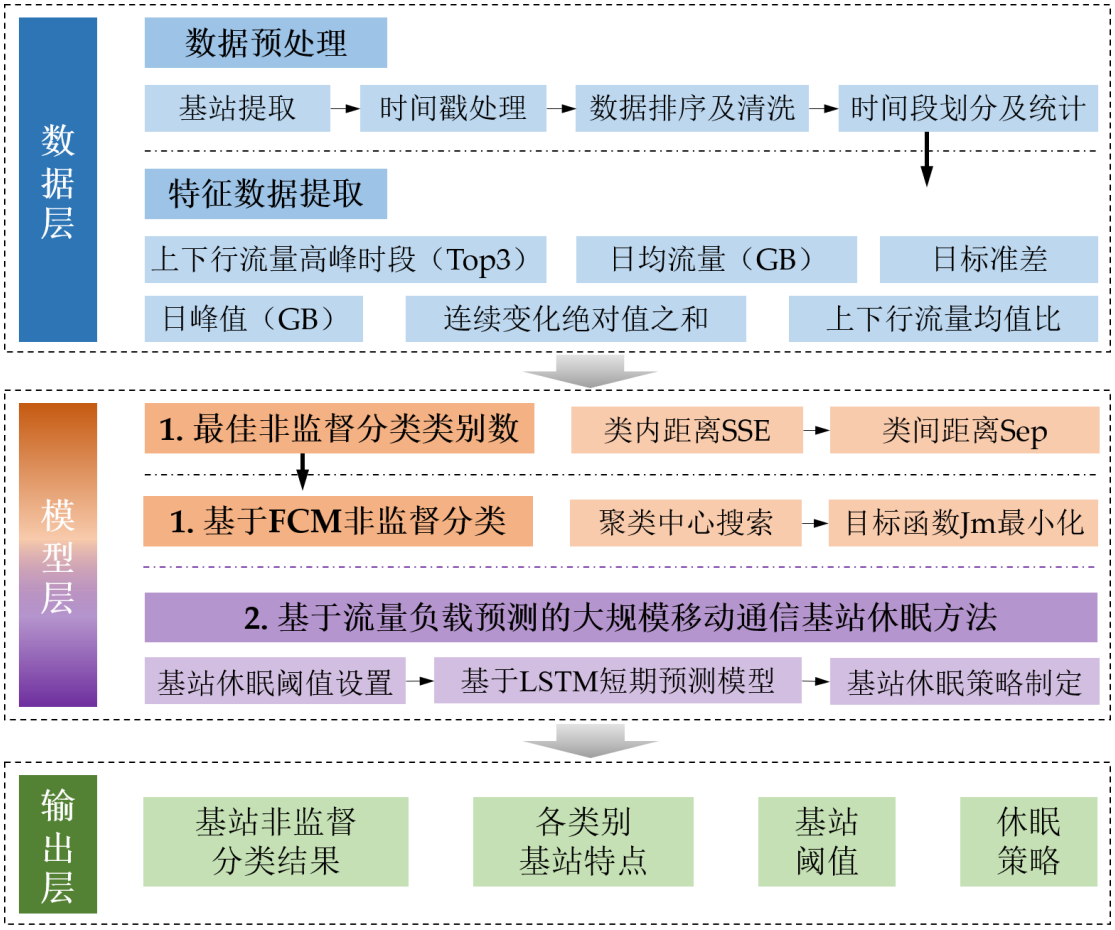


图 1. 大规模移动通信基站时序多特征非监督分类和基于流量负载预测的基站休眠方法框架图

3.1 大规模移动通信基站时序多特征非监督分类

3.1.1 数据预处理

由于现有历史数据存在样本量大，数据缺失多，时间戳混乱等问题，在建立模型前需要对原始数据进行预处理。预处理主要包含基站分割、时间戳处理、数据排序及清洗和统计四个步骤。**基站提取：**根据小区编号提取不同基站的历史流量数据并单独存档，以提高数据读取效率及降低计算机资源。**时间戳处理：**根据原始时间相关字段进行标准化，构建标准格式的时间戳字段，以方便时间序列的处理。**数据排序及清洗：**根据标准时间戳重新对基站历史流量数据进行排序，同时清洗重复导出带来的冗余数据及异常数

据（例如负值）。**时间段划分及统计**：按照研究思路，我们按照工作日及节假日的各个小时区间统计相应业务流量，并再次清洗包含空缺值的小区基站数据，保留具有完整时序特征的基站记录。

3.1.2 特征数据提取方法

在数据预处理基础上，我们对各个小区基站的时序特征进行提取。分别提取工作日及节假日的上下行流量高峰时段（Top3）、日均流量（GB）、日标准差、日峰值（GB）、连续变化绝对值之和以及上下行流量均值比作为小区基站历史流量数据的时序特征。其计算方法如下公式（1）-（6）所示：

$$\text{高峰时段：}\ T_{top} = \arg\max(\mathrm{U}_t) \tag{1}$$

$$\text{日均流量：}\ avg = sum(\mathrm{U}_t)/24 \tag{2}$$

$$\text{日标准差：}\ std = \sqrt{\frac{1}{24}\sum_{t=1}^{24}(\mathrm{U}_t - \mathrm{avg})^2} \tag{3}$$

$$\text{日峰值：}\ \mathrm{U}_{\max} = \max(\mathrm{U}_t) \tag{4}$$

$$\text{连续变化绝对值之和：}\ sum\_of\_change = \sum_{t=1}^{23}|\mathrm{U}_{t+1} - \mathrm{U}_t| \tag{5}$$

$$\text{上下行流量均值比：}\ D = avg_{up}/avg_{down} \tag{6}$$

最终根据按天的业务流量数据统计得到 30 个时序特征指标，用于后续建模。

3.1.3 基于改进 FCM 的自动非监督分类方法与类别特点分析方法

考虑现有基站数据无监督信息，本文引入 FCM 方法实现大规模移动基站的非监督分类。其次，由于 FCM 需人工预先输入特征数据聚类数目，其缺乏自适应能力，因此本文拟将其进行改进。在进行分监督分类之前，设计最佳聚类数目 K 的自动选取方法，而后再进行基站特征数据的非监督分类。因此，本方法模型主要包括：最佳聚类数目 K 的自动确定方法和基站非监督分类方法。

（1）最佳聚类数目 K 的自动确定方法

我们希望能从数据自身出发去确定真实的聚类数，也就是对数据而言的最佳聚类数，手肘法的思想在此处被运用。手肘法的核心指标是 SSE(sum of the squared errors，误差平方和)，如公式（7）所示。其核心思想是：随着聚类数 K 的增大，样本划分会更加精细，每个簇的聚合程度会逐渐提高，那么误差平方和 SSE 自然会逐渐变小。并且，当 K 小于真实聚类数时，由于 K 的增大会大幅增加每个簇的聚合程度，故 SSE 的下降幅度

会很大，而当 K 到达真实聚类数时，再增加 K 所得到的聚合程度回报会迅速变小，所以 SSE 的下降幅度会骤减，然后随着 K 值的继续增大而趋于平缓，也就是说 SSE 和 K 的关系图是一个手肘的形状，而这个肘部对应的 K 值就是数据的真实聚类数

$$SSE = \sum_{i=1}^{K}\sum_{j=1}^{N} \| x_j - U_i \|^2 \qquad (7)$$

其中，N 表示样本点数目，x_j 表示第 j 个样本点，U_i 表示第 i 个聚类中心。

然而，*SSE* 只能表示聚类过程中类内的距离，类间的距离（*Sep*）是作为类别可分性的重要指标之一，因此，在本研究中类间距离的计算也被考虑在内，计算如公式（8），其中U_k表示第 k 个聚类中心。

$$Sep = \min_{i \neq k} \| U_i - U_k \|^2 \qquad (8)$$

综合研判上述两个指标，自动确定特征数据该分多少类。

（2）基站非监督分类方法

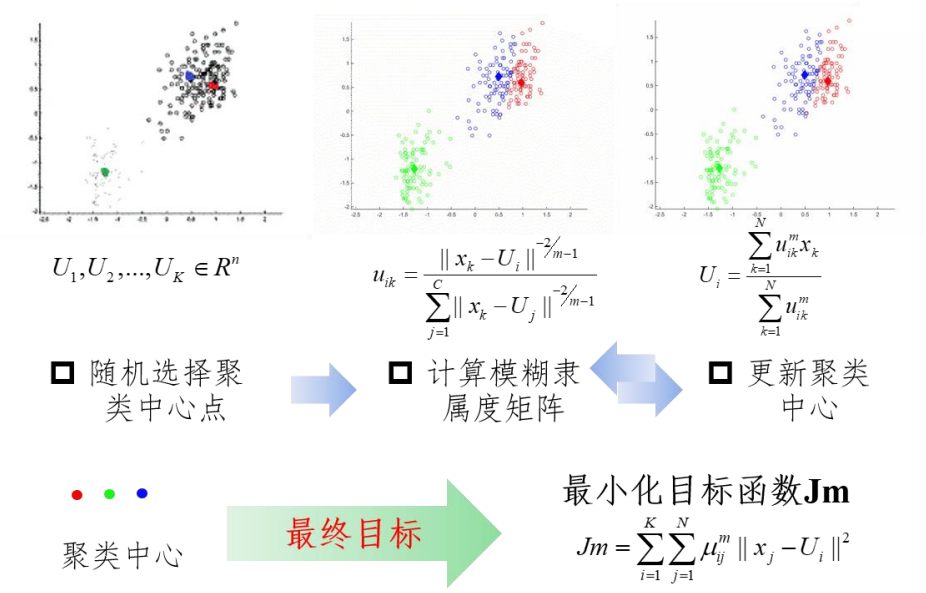


图 2. FCM 模型框架图

不同于硬聚类 K-means，其对于一个样本点的判别结果非 0 即 1，事实上难以符合实际的模糊世界，引入模糊聚类概念，即 FCM 非监督分类方法，并将第一部分的自动确定聚类数目加入改进已有的 FCM 非监督分类方法。其中，对于 FCM 方法而言，是一种基于划分的聚类算法，它的思想就是使得被划分到同一簇的对象之间相似度最大，而不同簇之间的相似度最小。模糊 C 均值算法是普通 C 均值算法的改进，普通 C 均值算法对于数据的划分是硬性的，而 FCM 则是一种柔性的模糊划分。FCM 用隶属度确定每个数据点属于某个聚类的程度的一种聚类算法。1973 年，Bezdek 提出了该算法，作为

早期硬 C 均值聚类（HCM）方法的一种改进。其主要流程图如图 2 所示。

根据第一部分基于类内和类间距离指标确定特征数据的最佳非监督分类数目，输入到 FCM 聚类过程当中。第一步，随机产生相应数目的聚类中心，其次计算模糊隶属度矩阵，如公式（9）所示。最后，继续更新迭代聚类中心，如公式（10）所示。因此，FCM 非监督分类的最终目标是使得聚类目标函数 *Jm* 最小化，计算如公式（11）。

$$\mu_{ij}=\frac{1}{\sum_{k=1}^{K}\left(\frac{1/\|x_j-U_i\|}{1/\|x_j-U_k\|}\right)^{2/(m-1)}},\quad \sum_{i=1}^{K}\mu_{ij}=1,\quad j=1,...,N \tag{9}$$

$$U_i=\frac{\sum_{j=1}^{N}u_{ij}^{m}x_j}{\sum_{j=1}^{N}u_{ij}^{m}} \tag{10}$$

$$Jm=\sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^{m}\|x_j-U_i\|^2 \tag{11}$$

其中，μ_{ij}为模糊隶属度矩阵，其满足和为 1 约束。算法的输出是 K 个聚类中心点向量和 K*N 的一个模糊划分矩阵，这个矩阵表示的是每个样本点属于每个类的隶属度。根据这个划分矩阵按照模糊集合中的最大隶属原则就能够确定每个样本点归为哪个类。聚类中心表示的是每个类的平均特征，可以认为是这个类的代表点。

在进行非监督分类的基础上，为说明每一类基站的特点，我们需要计算每一类中各特征数据的众数和中位数（不使用均值，避免噪声样本点的影响），然后分析各类别之间的众数和中位数的差异，反推得到每一类基站的特点。其中，高峰时段计算的是众数，其他特征计算的为中位数。

3.2 基于流量负载预测的大规模移动通信基站休眠方法

为了满足日益增长的数据业务需求，提高网络容量，宏基站与低功耗的小基站进行协作。部门的异构蜂窝网络已成为主流。无线通信网络的设计与规划是以最大负载为基础的，但难度较大。无线网络的流量是随着时间而变化的。当用户流量需求处于较低的过剩容量时，就会出现异构蜂窝网络。如果所有基站在低负荷期间仍处于活动状态，将会导致无线通信网络不稳定。基站睡眠技术的应用已经成为解决能耗问题的有效途径。通过对网络未来一段时间的分析。准确的流量预测可以更好地管理和分配网络资源，保证用户服务质量，提高资源利用效率。因此，本小节主要探讨基于流量负载预测的大规模移动通信基站休眠方法，其中主要包括：基站休眠阈值的设置、基于长短期记忆网络

（LSTM）的短期预测模型和基站休眠策略的制定。

3.2.1 基站休眠阈值的设置

针对网络负载的时间波动性，通过网络负载预测数据与实时网络负载情况进行联合分析，对网络中的皮基站（Pico BS）进行休眠与唤醒操作，以此达到减少系统能耗，提升网络能效的目标。假设网络架构中有集中控制器，可以采集基站的负载，工作状态，UE 位置等信息，具有分析预测等数据处理功能，可以对基站进行休眠和唤醒。 通过将一天内网络负荷的平均变化除以 X 个时间点 Q，可以形成 X-1 个时间段 T。在每个时间点 Q，设置下一个时间段 T 中网络中所有 PicoBS 的状态，并假定执行时间可以忽略不计。根据网络负载预测的情况，在每个操作设置时间点 Q 对下一时间段 T 休眠基站数量进行设定。设整个网络的归一化网络负载 $LT \in [0,1]$，通过设置三个归一化负载限制值 β_1，β_2，β_3，将网络负载水平划分为四个等级，划分情况如表 2 所示。

表 2. 基站通信网络负载水平等级划分

基站通信网络负载水平	归一化负载值
超低负载水平	$LT \in [0, \beta_1]$
低负载水平	$LT \in [\beta_1, \beta_2]$
中负载水平	$LT \in [\beta_2, \beta_3]$
高负载水平	$LT \in [\beta_3, 1]$

3.2.2 基于长短期记忆网络（LSTM）的短期预测模型

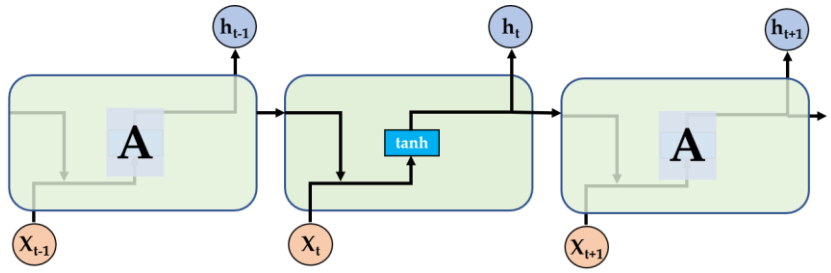


图 3. 标准 RNN 中的重复模块包含单个层

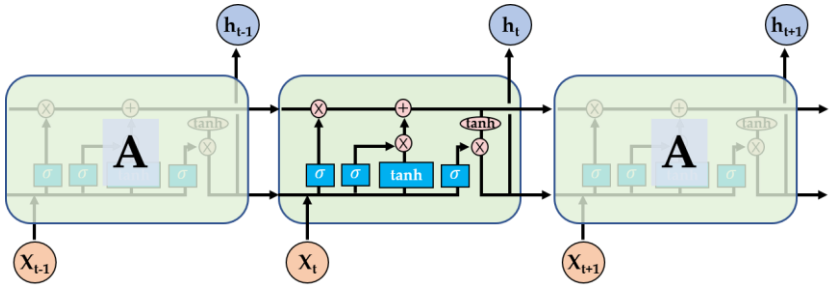


图 4. LSTM 中的重复模块包含四个交互层

LSTM 是一种特殊的循环神经网络（RNN），能够学习长期依赖关系。它们由霍赫里特和施密德胡伯（1997 年）提出[13]，并在以下工作中被许多人提炼和普及。它们在各种问题上工作得非常好，现在被广泛使用。其被显式设计，以避免长期依赖性问题，长时间记住信息实际上是他们的默认行为，因此，可以认为在具有周期性规律的基站数据预测方面有较大用处。所有循环神经网络都有神经网络重复模块链的形式。在标准 RNN 中，此重复模块将具有非常简单的结构，例如单个 tanh 层，如下图 4 所示。LSTM 也有这种链式结构，但重复模块具有不同的结构，而不是有一个神经网络层，四个以一种非常特殊的方式交互。

循环神经网络与多层感知机不同的地方是上个神经元的输出是下一个神经元的输入，但是随着重复模块越来越多，权重 W 若很小，则使得第一个神经元的信息就会在权重 W 相乘之后出现特别小的值，因会使得第一个模块的信息失真，同理若权重 W 很大，则会使得第一个模块的输出覆盖住所有的信息，也会使得信息失真。而 LSTM 模型与普通的 RNN 不同之处就是重复模块加入门的概念，门可以对上一个输出进行选择是否忘记，从数学上听说反向传播导数是一个常数，所以能够保证每个模块的信息能够被记住。

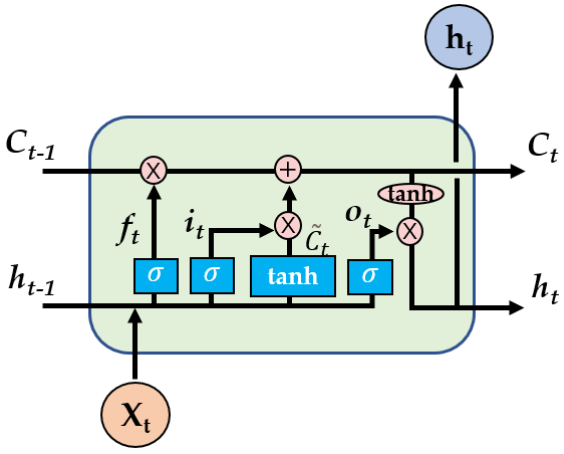


图 5. LSTM 单元格内关键计算

LSTM 的第一步是决定我们将从单元格状态中扔掉哪些信息。这个决定是由一个叫做"忘记门层"的西格莫德层做出的。它看着 h_t=1 和 X_t，并输出一个数字 0 和 1 单元格状态中每个数字 C_t=1.A1 表示"完全保留此"，而 0 表示"完全摆脱此"，如图 5 中所示，关键计算的公式（12）至（16）。

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (12)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (13)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (14)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (15)$$

$$h_t = o_t \times \tanh(C_t) \quad (16)$$

3.2.3 基站休眠策略的制定

在获取当前时段网络全部用户的流量状态及当前的休眠比例参数 z_q，依照下一时段 T 的负载水平 z_p，确定下一时段休眠比例参数的大小。如公式（17），如果 z_q 大于 z_p，则说明下一时间段负载水平比当前有所提升，需要唤醒部分休眠基站；如果 z_q 小于 z_p，则说明下一时间段负载水平比当前有所下降，需要休眠更多的基站；如果 z_q 等于 z_p，则说明下一时间段负载水平比当前相同，应保持当前休眠基站数量。

$$\text{基站休眠策略}\begin{cases}\text{唤醒部分休眠基站（}z_q\text{大于}z_p\text{）}\\ \text{休眠部分基站（}z_q\text{小于}z_p\text{）}\\ \text{保持不变（}z_q\text{等于}z_p\text{）}\end{cases} \quad (17)$$

四、结果与分析

4.1 大规模移动通信基站多特征分类结果与特点分析

4.1.1 数据预处理和特征数据提取结果和分析

表 3. 历史流量数据特征表

特征名称		特征值
记录数		144138200
小区（基站）数		132279
初始时间		2018/3/19 0:00:00
结束时间		2018/4/18 23:00:00
上行业务量 GB	平均值	0.05049276
	标准差	0.09017596
	最小值	0
	最大值	6.283223
下行业务量 GB	平均值	0.3537783
	标准差	0.6388047
	最小值	0
	最大值	39.313

（1）基站数据预处理

历史数据：本次比赛给出 132279 个小区数据，共计 144138200 条历史记录，每条记录包含“日期”，“时间”，“小区编号”，“上行业务量 GB”，“下行业务量 GB”5 个字段。所有记录被统一存放于 csv 文件中，历史数据的整体特征如上表 3 所示。

由于比赛给的历史数据存在样本量大，数据缺失多，时间戳混乱等问题，在建立模型前需要对原始数据进行预处理。预处理主要包含基站分割、时间戳处理、数据排序及清洗和统计四个步骤。

（a）基站提取：历史数据文件过大（9GB），导致计算机读取文件时间较长且难以在小内存的环境下进行处理。在假设中提到，各基站保持独立且不受其他基站影响，我们认为每个基站可以单独被提取进行分析，因此根据小区编号字段分离出各基站对应流量记录，每个小区保存成单独文件。

（b）时间戳处理：经由目标小区提取后，每个文件中保存单独小区流量记录，该记录字段与历史数据中保持一致。数据检查后发现，部分原始日期信息导出不完整，如 no.186 小区 4 月 1 号的日期标注为“018-04-01”，这类信息难以被格式化成系统有效的日期格式，因此需要作出预处理。我们将该日期和时间字段合并，构建标准化时间戳，如“2018/3/1　1:00:00”，并以“Time”字段进行记录。

（c）数据排序及清洗：经检查后发现业务流量数据并非完全按照时间顺序排列，导致历史数据无法直接用于构建标准时间序列，因此我们根据上述创建的“Time”字段进一步对历史记录进行排序。此外，考虑到数据库的重复入库及导出操作，我们针对各基站记录进行数据清洗，清洗时间戳相同的冗余数据以及显著出现错误的异常值。

（d）统计：按照上述问题 1 的研究思路，我们需要获取工作日及节假日各个小时区间的业务流量信息。我们统计了 2018 年 3 月及 4 月的国定节假日作为筛选要素，根据记录 Time 字段确定其该记录是否属于节假日，同时确定该记录位于一天中的时间区间。我们将工作日 24 时段及节假日 24 时段的记录进行了汇总并计算其业务流量的平均值。由于该部分保留了按天统计信息，缺失了月份信息，因此我们同时统计了历史数据中该月的流量最高的日期加入时序特征。最后，根据统计获得的工作日及节假日时段流量信息，再次清洗具有空缺值的小区基站数据，保留具有完整时序特征的基站记录。

通过步骤（a）-（d），去除具有空缺值得基站数据，数据样本数目从 132279 条到 118756 条。

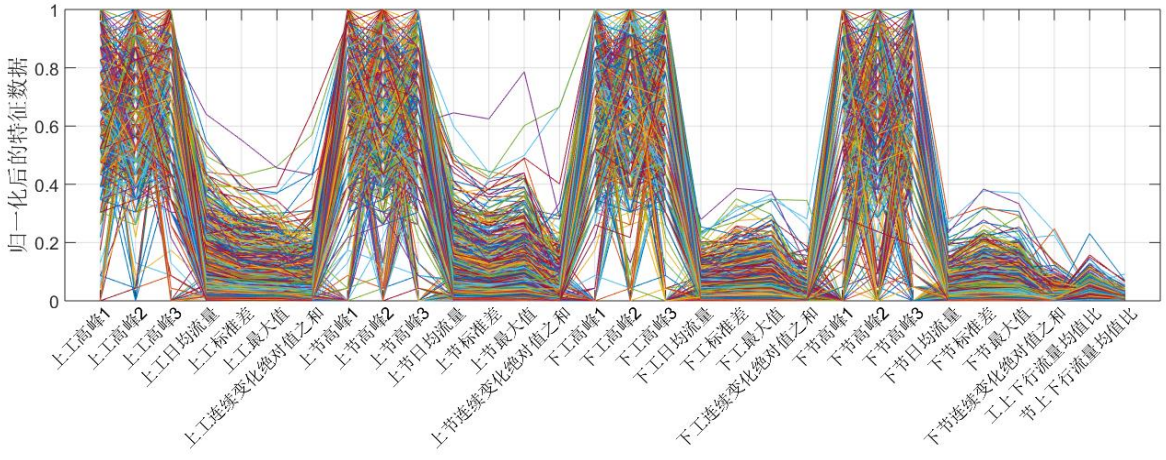
(2) 特征数据提取结果和分析

根据方法模型中的特征数据提取方法，我们对各个小区基站的时序特征进行提取。

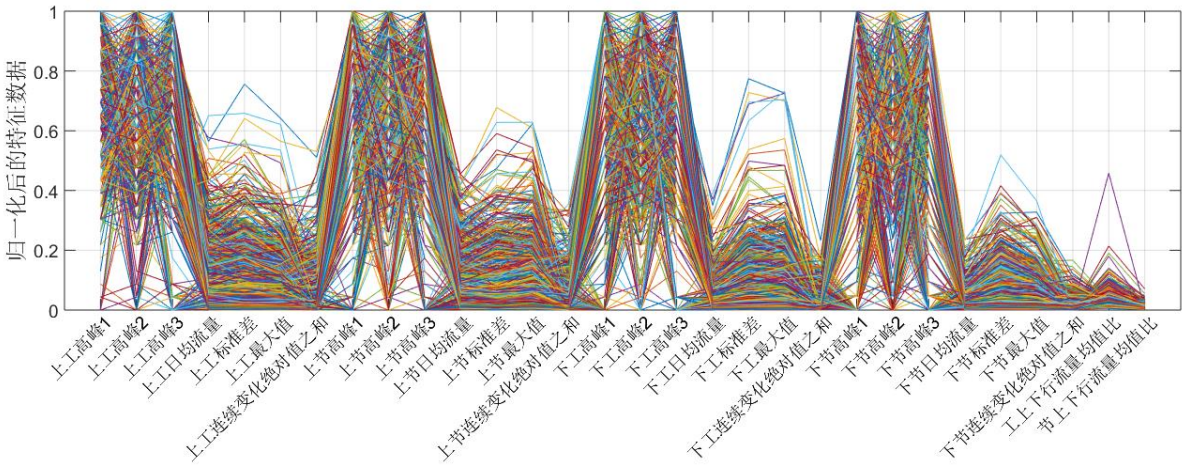
分别提取工作日及节假日的上下行流量高峰时段（Top3）、日均流量（GB）、日标准差、日峰值（GB）、连续变化绝对值之和以及上下行流量均值比作为小区基站历史流量数据的时序特征，共提取到 30 维的数据。如表 4 所示，其展示的是 1 号基站进行特征提取后的特征数据，将用以后续非监督分类。从表 4 中可以发现，分节假日和工作日进行统计的必要性，如工作日基站下行的日均流量为 2.74623GB，而节假日基站下行得日均流量为 3.09737GB，这也正好符合节假日所用流量会更多得特点，便可推测 1 号基站位于居民区或者商业区的概率大于办公区，因此，我们所提取的特征数据在理论上分析是有利于后续的非监督分类过程的。

表 4. 1 号基站提取后的特征数据

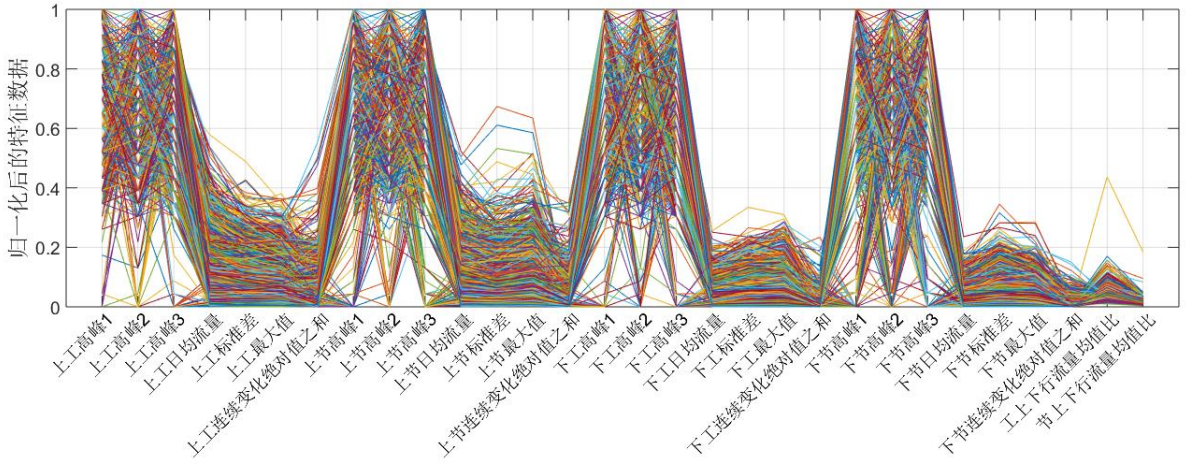
	特征		基站上行	基站下行
工作日	高峰时段	No.1	21	22
		No.2	22	21
		No.3	20	23
	日均流量（GB）		0.33399	2.74623
	日标准差		0.18621	1.57896
	日峰值（GB）		0.70973	6.19034
	连续变化绝对值之和		5.00343	19.3131
	上下行流量均值比		0.12162	
节假日	高峰时段	No.1	21	21
		No.2	22	20
		No.3	20	19
	日均流量（GB）		0.36012	3.09737
	日标准差		0.17770	1.50620
	日峰值（GB）		0.61455	5.39651
	连续变化绝对值之和		2.31061	11.72395
	上下行流量均值比		0.11627	



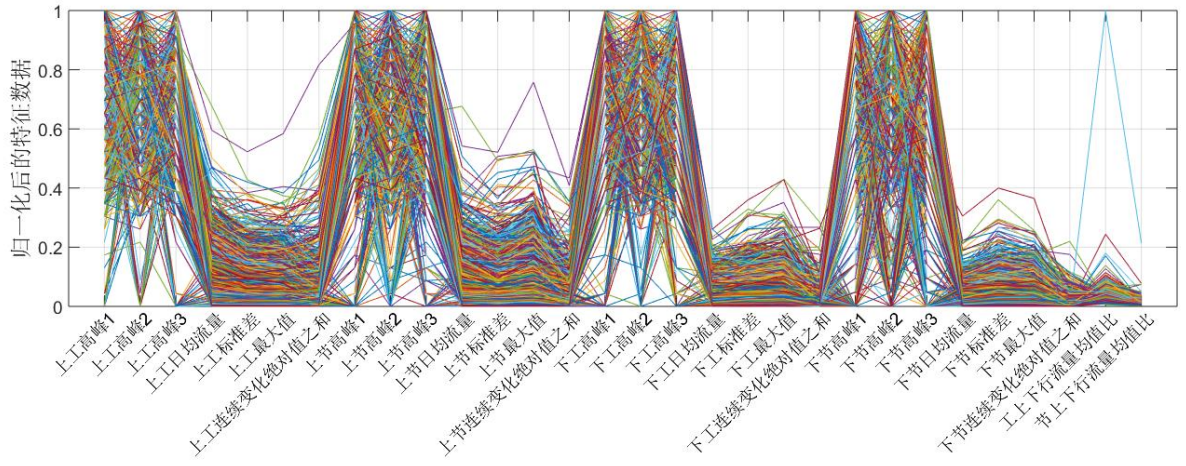
（a）排序在 1-5000 的归一化基站特征数据



（b）排序在 10001-15000 的归一化基站特征数据



（c）排序在 50001-55000 的归一化基站特征数据



（d）排序在 90001-95000 的归一化基站特征数据

图 6. 归一化后的部分特征数据展示

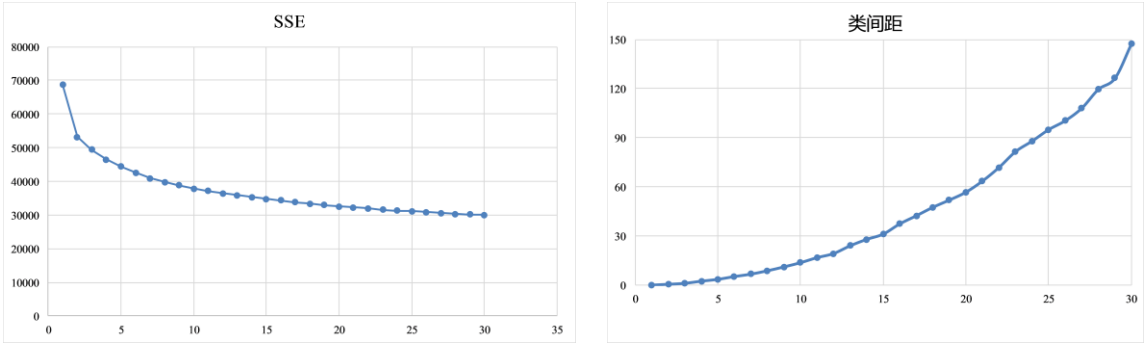
其次，为有效进行后续的非监督分类，考虑所提取的特征数目并非都处于统一量纲，因此，归一化操作被引入，将所有数据按列归一化到统一量纲之下，基站数据按照基站

序号进行排列。此外，将归一化后的高维特征数据进行平行坐标轴展示，如图 6 所示，分别展示了 1-5000，10001-15000，50001-55000 和 90001-95000 的特征数据。

4.1.2 基站非监督分类结果分析

（1）综合分析最非监督分类类别数目

根据最佳聚类数目 K 的自动确定方法，计算特征数据的类内距离和类间距离，如图 7 所示，可以发现类内距离在 K 值为 2 是有一个突变，但其并非理想的非监督分类数目，结合类间距离，我们可以发现在 K 值为 12 时，其后有一个上升的突变。因此，综合分析类内距离和类间距离，我们选择 K 值为 12 作为最优的非监督分类类别数，将其输入到后续的 FCM 多时序特征数据进行非监督分类。



（a）类内距离 SSE　　（b）类间距离

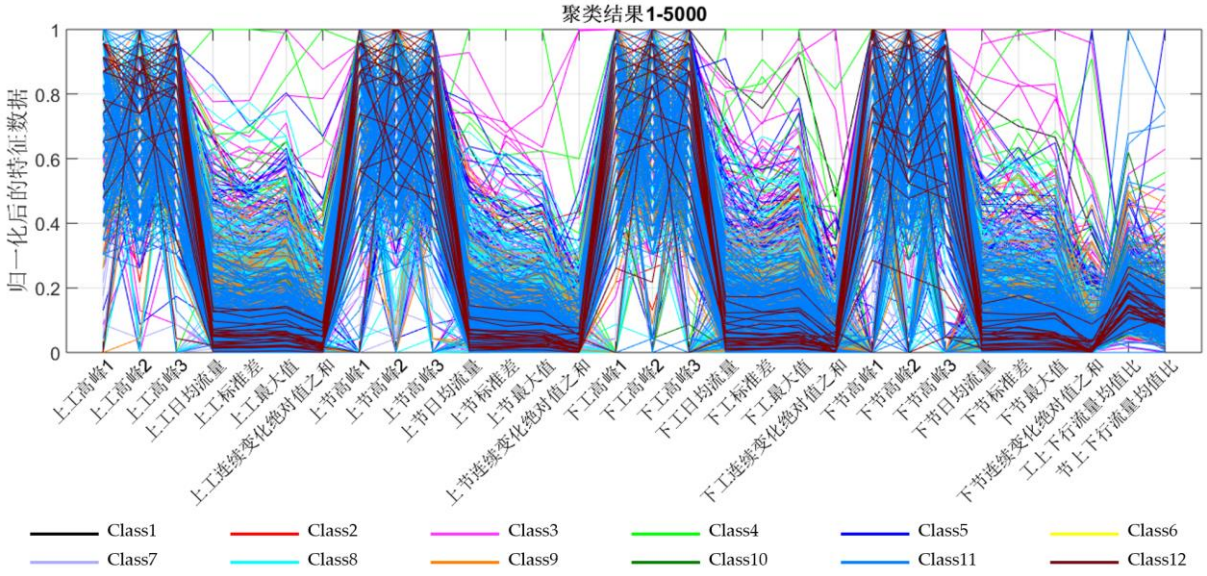
图 7. 类内距离和类间距离指标

（2）FCM 非监督分类结果与分析

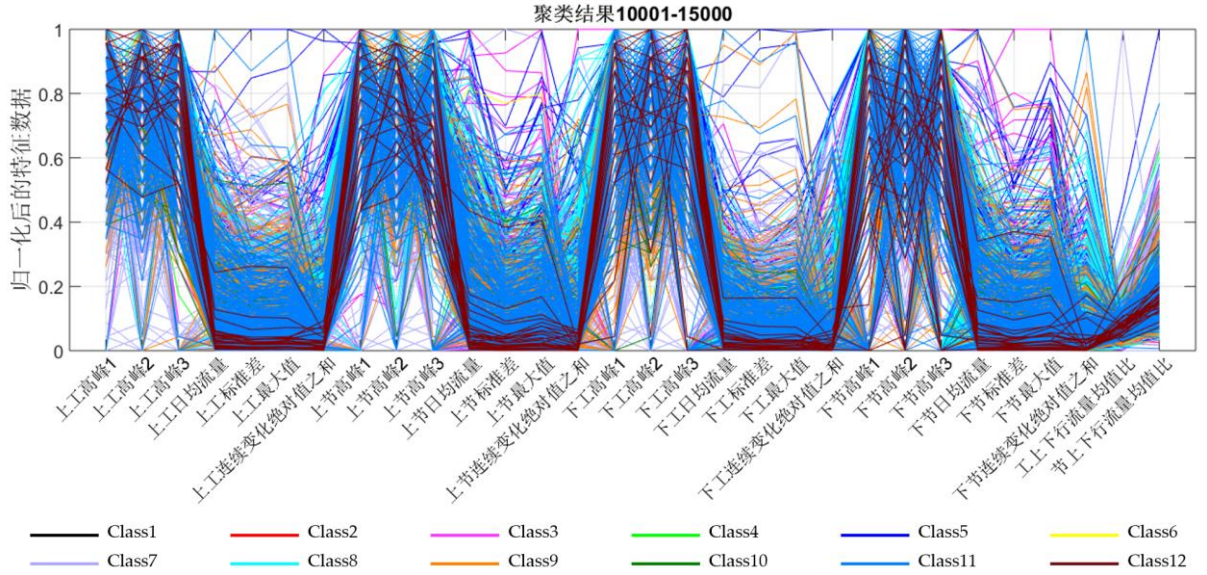
根据最 FCM 非监督分类方法模型，输入第一部分确定的最优分类数目 12，将按列归一化后的大小为 118756×30 的特征数据进行非监督分类，各类别所分类的数据和所占比率如表 5 所示。从表 5 中，可以看出，第 7 类所占比率最大，其次为第 3 类，占比最小的类别为第一类，仅为 0.05%。另外，考虑高维数据以及样本点过多，我们依旧展示对应于之前四段数据（剔除缺失值的基站号按顺序排序为 1-5000，10001-15000，50001-55000 和 90001-95000）的非监督分类结果，如图 8 所示。

表 5. 各类别所分类的数据和所占比率

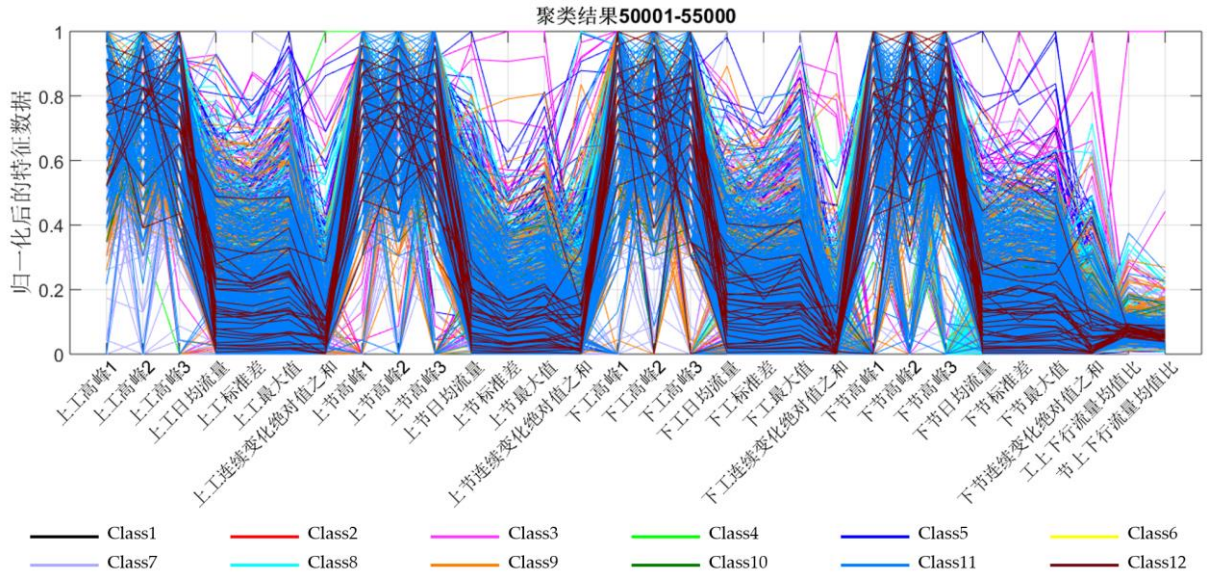
类别	1	2	3	4	5	6
数目	57	959	19308	1261	17222	1353
所占比率	0.05%	0.81%	16.26%	1.06%	14.50%	1.14%
类别	7	8	9	10	11	12
数目	33651	18626	12393	808	12501	617
所占比率	28.34%	15.68%	10.44%	0.68%	10.53%	0.52%



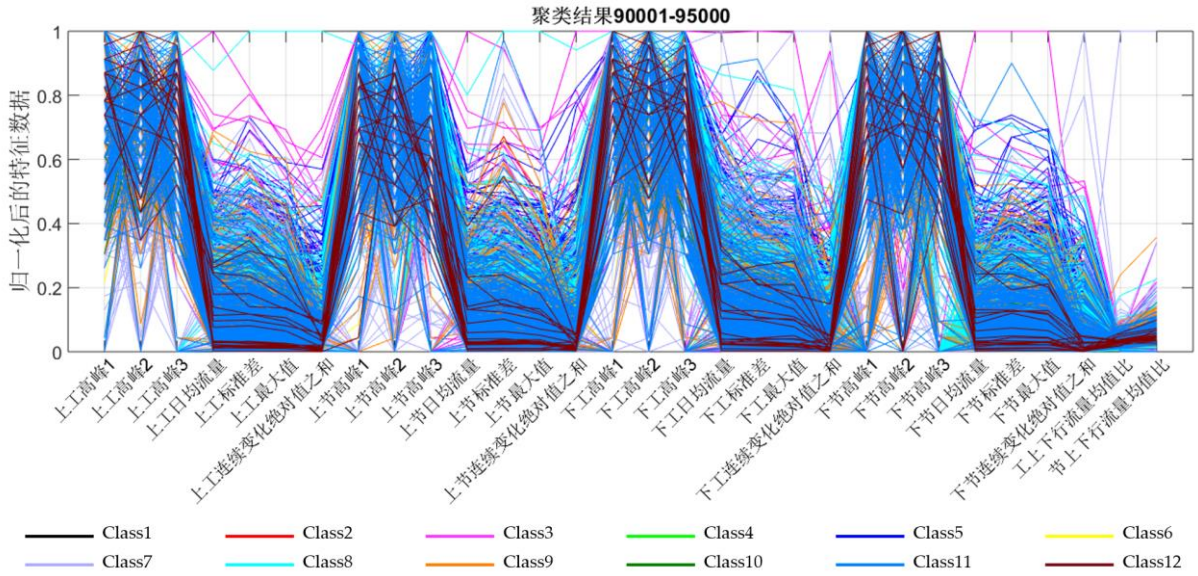
（a）排序在 1-5000 的归一化基站特征数据非监督分类结果



（b）排序在 10001-15000 的归一化基站特征数据非监督分类结果



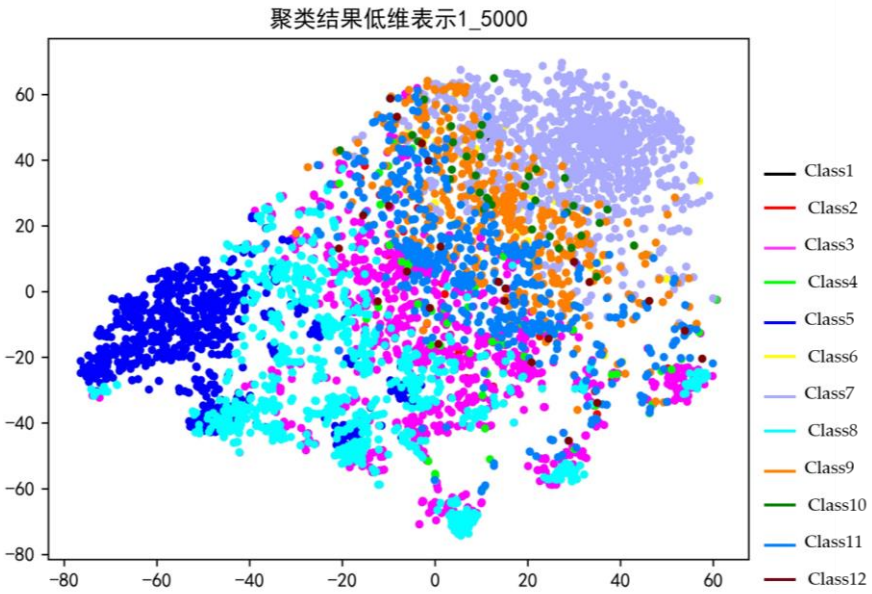
（c）排序在 50001-55000 的归一化基站特征数据非监督分类结果



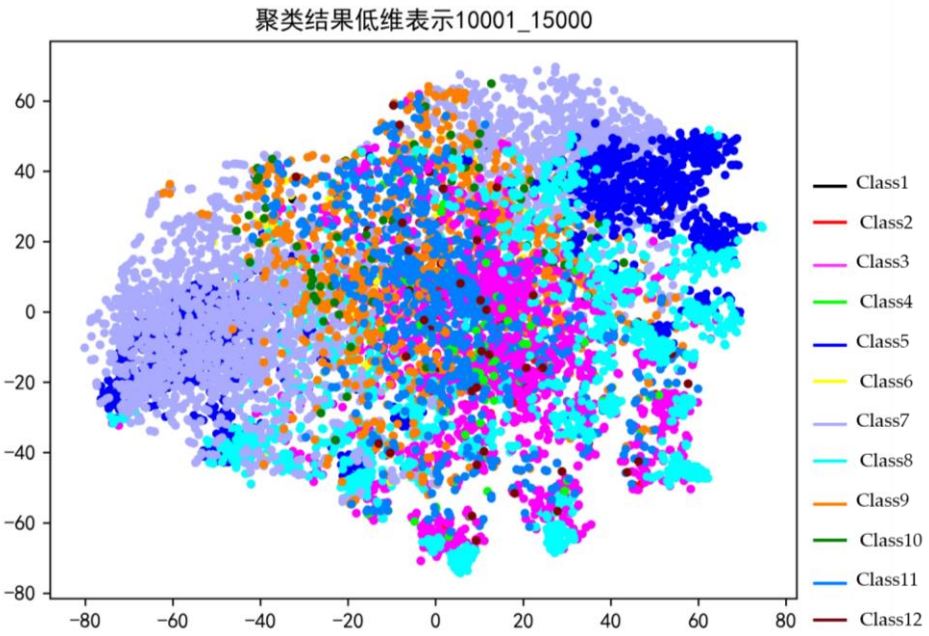
（d）排序在 90001-95000 的归一化基站特征数据非监督分类结果

图 8. 部分特征数据（剔除缺失值的基站号按顺序排序为 1-5000，10001-15000，50001-55000 和 90001-95000）的非监督分类结果

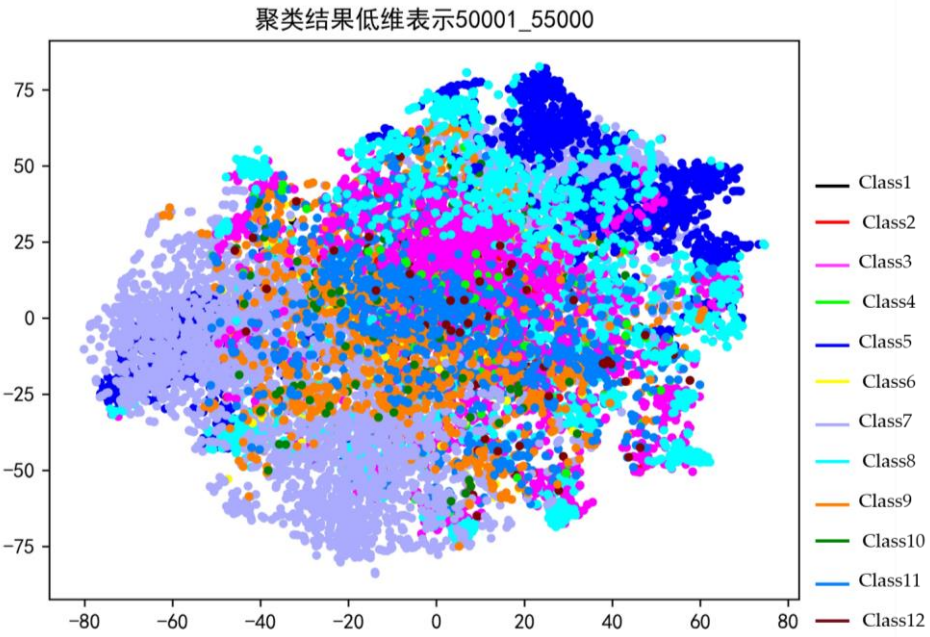
通过图 8，可以看出，很多不同类别的特征数据由于重叠难以看清。因此，我们考虑降维后进行数据非监督分类后的展示。考虑 t-Distributed Stochastic Neighbor Embedding （t-SNE）是一种降维技术，用于在二维或三维的低维空间中表示高维数据集，从而使其可视化。与其他降维算法(如 PCA)相比，t-SNE 创建了一个缩小的特征空间，相似的样本由附近的点建模，不相似的样本由高概率的远点建模。降维后进行展示的结果如图 9 所示。



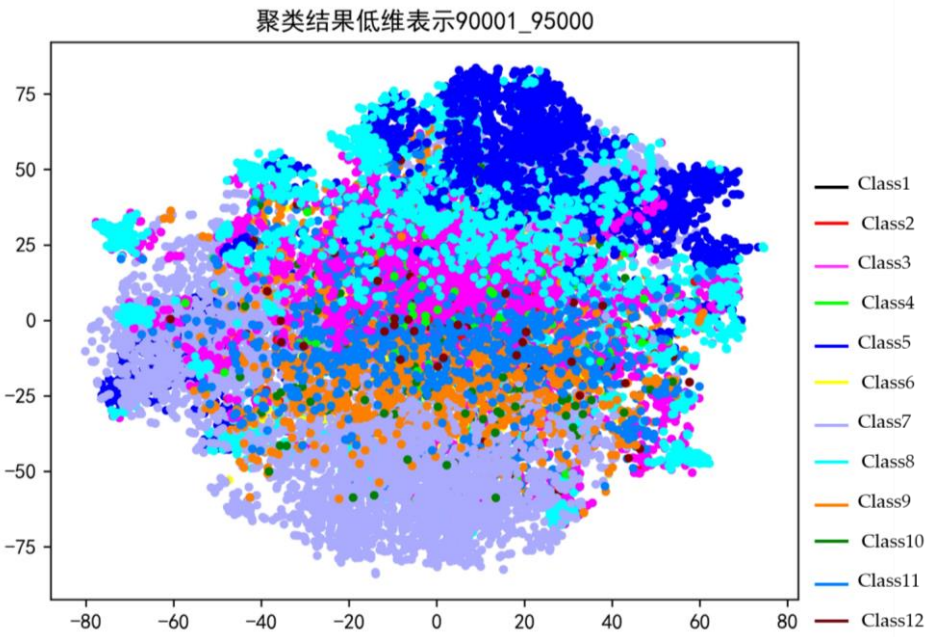
（a）排序在 1-5000 的归一化基站特征数据非监督分类结果（t-SNE）



（b）排序在 10001-15000 的归一化基站特征数据非监督分类结果（t-SNE）



（c）排序在 50001-55000 的归一化基站特征数据非监督分类结果（t-SNE）



（d）排序在 90001-95000 的归一化基站特征数据非监督分类结果（t-SNE）

图 9. 部分特征数据（剔除缺失值的基站号按顺序排序为 1-5000，10001-15000，50001-55000 和 90001-95000）的非监督分类结果（t-SNE）

4.1.3 基站分类类别特点分析

为进一步分析各类别小区基站的流量特点，本文针对各类别小区的历史流量特征进行统计。针对高峰时段特征，采用各类别内基站高峰时段的众数作为分析指标；针对其他浮点数型统计特征，本文采用各类别内基站相应时序特征的中位数作为分析指标，中位数相比于平均值对异常值的抗敏感性更强。对应前面提到的 30 种特征构建 f1-f30 共 30 个分析指标以作进一步分析，如表 6 所示。本文对 12 类小区基站流量特点的分析如下：

C1：整体流量低，流量稳定，高工作日流量，高上下行流量比。C1 具有所有类别中最小的日流量均值，用户密度较小。C1 具有最小的上下行日流量标准差，其流量使用情况十分平稳。工作日上行日均流量同比节假日高出 11.2%，下行日均流量同比高出 32.4%，工作日流量使用显著高于节假日。此外，尽管 C1 的上传量均值较小，但其具有最高的上下行流量均值比，该小区基站整体流量使用较小，但有较为明显的上传特征。因此我们推测 **C1 基站附多为工作区域**。从峰值情况来看，工作日流量高峰期普遍出现于 12 点、15 点等工作时间段，符合工作区流量使用特征。

C2：节假日高峰时段提前，高节假日流量，节假日波动明显。C2 工作日高峰时段集中于 18 点-19 点，而节假日高峰显著提前到 16-18 点。节假日上行日均流量同比工作日高出 9.0%，下行日均流量同比高出 10.2%，节假日上下行流量峰值均高于工作日流量

峰值，节假日流量使用显著高于工作日。此外，在所有类别中 C2 的工作日上行流量标准差排名第 8，下行流量排名第 6；节假日上行流量标准差排名第 3，下行流量排名第 3，节假日的波动性更为显著。因此我们推测 **C2 基站附近多为餐饮购物等商业区域**。

C3：高峰时段稳定，高节假日流量，节假日波动明显。C3 无论工作日还是节假日高峰时段都集中于 21-22 点。节假日上行日均流量同比工作日高出 6.4%，下行日均流量同比高出 7.3%，节假日流量使用高于工作日。此外，C3 的节假日流量使用水平整体高于工作日，但工作日上行流量连续变化绝对值之和高达 0.268，下行流量连续变化绝对值之和高达 0.555，远高于节假日上下行流量连续变化绝对值之和 0.158 和 0.145，节假日的流量使用情况更加平稳。因此我们推测 **C3 基站附近多为居住区域**，21-22 点的高峰时段正好符合居家特征。

C4：节假日高峰时段延迟，高节假日流量。C4 工作日高峰时段集中于 18 点-19 点，而节假日高峰显著延迟到 21-22 点，节假日晚间流量使用较为活跃。节假日上行日均流量同比工作日高出 1.7%，下行日均流量同比高出 6.7%，节假日上下行流量峰值与工作日流量峰值十分接近，节假日流量使用略高于工作日。C4 整体与 C2 类似，但节假日的高峰时段相对延迟。因此我们推测 **C4 附近多为混合区域。**

C5：整体流量大，高峰时段稳定，高节假日流量，低上下行流量均值比。C5 具有所有分类中最大的日均上下行日均流量、流量峰值及流量标准差。C5 整体流量特征与 C3 类似，具有稳定的流量高峰时段（20-21 点）；节假日上行日均流量同比工作日高出 7.5%，下行日均流量同比高出 6.7%，节假日流量使用高于工作日。由于整体流量水平较高，流量波动也呈现出明显的工作日特征，工作日流量连续变化绝对值之和远高于节假日。此外，C5 具有所有分类中最低的上下行流量均值比，该区域主要采用流量下载模式而非上传模式，这更符合居家流量使用的特征。因此我们推测 **C5 基站附近多为用户密集的大规模居住区域**。

C6：高工作日流量，工作日与节假日流量波动差异大，高上下行流量均值比。C6 高峰时段多数稳定在 12 点及 17 点左右。工作日上行日均流量同比节假日高出 16.5%，下行日均流量同比高出 24.0%，工作日流量使用显著高于节假日。C6 工作日上行流量标准差排名第 5，节假日排名第 8，工作日下行流量标准差排名第 4，节假日排名第 10，C6 的工作日流量波动显著大于节假日波动。此外，C6 在工作日及节假日具有排名第 3 的上下行流量均值比，该区域存在较多的上传任务。因此我们推测 **C6 附近多为工作区域，且用户密度大于 C1**。

C7：高工作日流量，整体流量稳定。C7 高峰时段多数集中于 12 点，且 C7 工作日

上行日均流量同比节假日高出 31.0%，下行日均流量同比高出 41.5%，工作日流量使用显著高于节假日。C7 整体流量特征（均值、峰值、标准差）与 C6 相似，但 C7 具有所有类别中最小的工作日上下行流量连续变化绝对值之和，在工作时间内具有最稳定的网络波动；节假日的上、下行流量连续变化绝对值之和分别位列第 2、第 1，该区域具有相当稳定的整体网络波动。因此我们推测 **C7 附近多为工作区域，且工作实时流量稳定。**

C8：整体流量大，高峰时段稳定，高节假日流量，低上下行流量均值比。C8 具有所有类别中排名第 2 的上下行流量均值，整体用户活跃。工作日和节假日高峰时段都集中于 20-21 点。节假日上行日均流量同比工作日高出 8.0%，下行日均流量同比高出 7.6%，节假日流量使用高于工作日。此外，C8 具有所有类别中最低的上下行流量均值比，流量使用过程中下载模式远多于上传模式，该模式符合居家流量使用特点。因此我们推测 **C8 基站附近多为居住区域，且用户密度大于 C3，居住区发展相对完善**。

C9：高峰时段稳定，高工作日流量。无论是工作日还是节假日，C9 的高峰时段稳定在 17 点左右。C9 的工作日上下行流量均值排在所有类别中的第 3，而节假日上下行流量均值排在第 7，工作日上行日均流量同比节假日高出 8.2%，下行日均流量同比高出 6.8%。因此我们推测 **C9 附近多为混合区域（工作区域、餐饮娱乐等商业区域）。**

C10：高峰时段波动，高工作日流量，高上下行流量均值比。C10 工作日流量时间序列有 12 点和 17 点两个峰值，节假日高峰时段集中在 16 点左右，两者特征具有明显差异。工作日上行日均流量同比节假日高出 15.9%，下行日均流量同比高出 25.2%，工作日流量使用显著高于节假日。与 C1 相似，C10 也具有极高的上下行流量均值比，上传流量使用率高，符合企业单位的流量使用特点。因此我们推测 **C10 附近多为工作区域，且用户密度较高**。

C11：高峰时段稳定，流量均值稳定。工作日和节假日的 C11 高峰时段均集中在 17-18 点，无明显差异。工作日上下行流量特征（均值、标准差、峰值）与节假日流量特征基本保持一致，无明显差异。因此我们推测 **C11 附近为混合区域。**

C12：整体流量低，节假日高峰时段延迟，高节假日流量。C12 上下行整体日均流量均位列最后几位，其中工作日上下行日均流量均位于所有类别的倒数第 2 位，突出表现工作日用户流量使用较少。C12 工作日流量高峰时段集中于 18 点，而节假日高峰流量呈现多波峰状态，其中上行流量最大值出现在 20 点下行流量最大值出现在 0 点，相较于工作日高峰时段有明显的滞后。此外，节假日上行日均流量同比工作日高出 3.1%，下行日均流量同比高出 6.4%，节假日流量使用高于工作日。因此我们推测 **C12 附近多为用户密度较低的居住区域**。

表 6. 计算每一类中各特征数据的众数和中位数

类别	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10
#1	12	17	14	0.0198	0.0148	0.0516	0.0911	17	15	14
#2	18	20	19	0.0277	0.0182	0.0646	0.2334	16	18	17
#3	20	21	20	0.0267	0.0170	0.0626	0.2675	22	21	21
#4	18	19	21	0.0269	0.0176	0.0617	0.2221	22	17	21
#5	21	21	20	0.0484	0.0325	0.1237	0.6729	20	21	21
#6	12	17	17	0.0284	0.0200	0.0663	0.1284	17	16	16
#7	12	10	11	0.0272	0.0208	0.0680	0.0806	10	10	11
#8	21	21	20	0.0331	0.0213	0.0798	0.4069	21	21	20
#9	18	17	17	0.0298	0.0201	0.0675	0.1510	17	16	16
#10	12	17	17	0.0279	0.0188	0.0611	0.1230	17	16	16
#11	18	18	19	0.0282	0.0183	0.0637	0.1883	17	17	17
#12	18	18	19	0.0229	0.0147	0.0506	0.1946	20	19	16
类别	**f11**	**f12**	**f13**	**f14**	**f15**	**f16**	**f17**	**f18**	**f19**	**f20**
#1	0.0178	0.0122	0.0447	0.0498	12	15	10	0.1310	0.0955	0.3763
#2	0.0301	0.0195	0.0673	0.1442	18	18	19	0.1956	0.1189	0.4381
#3	0.0284	0.0179	0.0649	0.1575	22	21	20	0.1894	0.1120	0.4255
#4	0.0274	0.0177	0.0636	0.1513	18	19	19	0.1877	0.1136	0.4329
#5	0.0520	0.0326	0.1244	0.3675	21	21	20	0.3652	0.2160	0.8377
#6	0.0237	0.0163	0.0569	0.0700	12	17	17	0.1960	0.1282	0.4505
#7	0.0188	0.0141	0.0485	0.0435	12	13	11	0.1853	0.1350	0.4590
#8	0.0357	0.0217	0.0813	0.2415	22	21	21	0.2456	0.1447	0.5557
#9	0.0274	0.0188	0.0639	0.0829	18	18	17	0.2036	0.1286	0.4539
#10	0.0235	0.0161	0.0547	0.0653	12	18	17	0.1860	0.1174	0.4103
#11	0.0282	0.0189	0.0658	0.1098	18	18	19	0.1949	0.1182	0.4286
#12	0.0236	0.0156	0.0542	0.1154	18	21	19	0.1547	0.0920	0.3357
类别	**f21**	**f22**	**f23**	**f24**	**f25**	**f26**	**f27**	**f28**	**f29**	**f30**
#1	0.1992	16	15	12	0.0989	0.0671	0.2175	0.1716	0.1606	0.1513
#2	0.4848	16	16	14	0.2178	0.1287	0.4670	0.3390	0.1457	0.1444
#3	0.5550	21	21	20	0.2043	0.1174	0.4284	0.3863	0.1452	0.1423
#4	0.4865	0	21	21	0.2012	0.1161	0.4264	0.3457	0.1470	0.1451
#5	1.3666	21	20	19	0.3913	0.2157	0.8466	0.8316	0.1363	0.1356
#6	0.2865	12	15	16	0.1580	0.1022	0.3554	0.2156	0.1528	0.1480
#7	0.1889	12	15	11	0.1309	0.0924	0.3186	0.1350	0.1487	0.1428
#8	0.8697	21	20	20	0.2658	0.1458	0.5520	0.5608	0.1397	0.1379
#9	0.3118	17	16	16	0.1906	0.1189	0.4112	0.2145	0.1510	0.1465
#10	0.2503	15	16	14	0.1485	0.1028	0.3392	0.1955	0.1530	0.1490
#11	0.3810	17	17	16	0.1967	0.1201	0.4232	0.2688	0.1504	0.1454
#12	0.3801	0	16	16	0.1653	0.1040	0.3734	0.2670	0.1478	0.1435

注：f1-f30 分别为：上工高峰 1，上工高峰 2，上工高峰 3，上工日均流量，上工标准差，上工最大值，上工连续变化绝对值之和，上节高峰 1，上节高峰 2，上节高峰 3，上节日均流量，上节标准差，上节最大值，上节连续变化绝对值之和，下工高峰 1，下工高峰 2，下工高峰 3，下工日均流量，下工标准差，下工最大值，下工连续变化绝对值之和，下节高峰 1，下节高峰 2，下节高峰 3，下节日均流量，下节标准差，下节最大值，下节连续变化绝对值之和，工上下行流量均值比，节上下行流量均值比。

4.2 大规模移动通信基站休眠策略结果与分析

4.2.1 相关阈值参数设置

我们可以根据预测流量，对应处于每一个负载水平的时段，提前设置相应的休眠比例参数（z_1、z_2、z_3、z_4），其中比例参数分别代表超低负载水平、低负载水平、中负载水平和高负载水平，具体参数设置如表 7 所示。

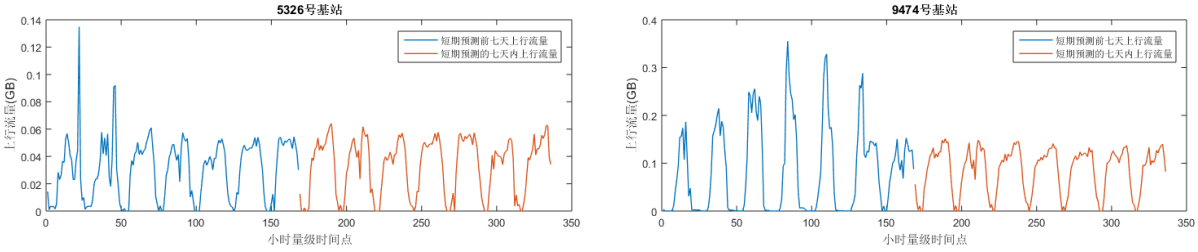
表 7. 基站休眠阈值方法参数设置

参数	数值	参数	数值	参数	数值
β_1	20%	z_1	80%	z_4	0
β_2	50%	z_2	60%	T	1
β_3	80%	z_3	30%	X	25

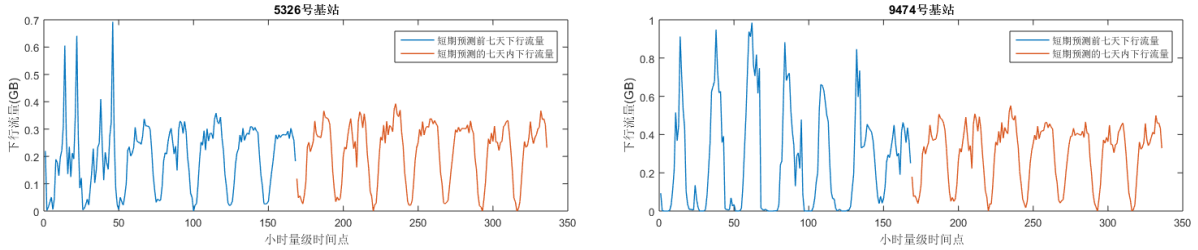
4.2.2 短期预测下的基站休眠策略结果与分析

（1）LSTM 短期预测结果

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后，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，给出了各个小区小时级上行和下行流量的预测模型，预测了这些小区后面一周(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。以部分基站（5326 号和 9474 号）上行和下行流量短期预测结果进行后续休眠策略制定，预测结果如下图 10 所示。



（a）上行流量



（b）下行流量

图 10. 5326 号和 9474 号基站上行/下行流量短期预测结果

（2）基于预测结果的基站休眠策略结果与分析

根据 5326 号和 9474 号基站上行/下行流量短期预测结果以及制定的休眠阈值，将短期预测后 24 小时内的数据进行下一步休眠结果分析。5326 号和 9474 号基站休眠/唤醒策略制定如下图所示，可以发现 5326 号基站和 9474 号基站存在几乎类似的休眠和唤醒策略，在凌晨时刻通常需要将基站进行休眠，7:00-21:00 时为唤醒状态，22:00-24:00 时为休眠状态。通过推算，5326 号和 9474 号基站可能处于商业区或者办公区，并且基于此休眠/唤醒策略，可有效降低近 37.5%的能耗。

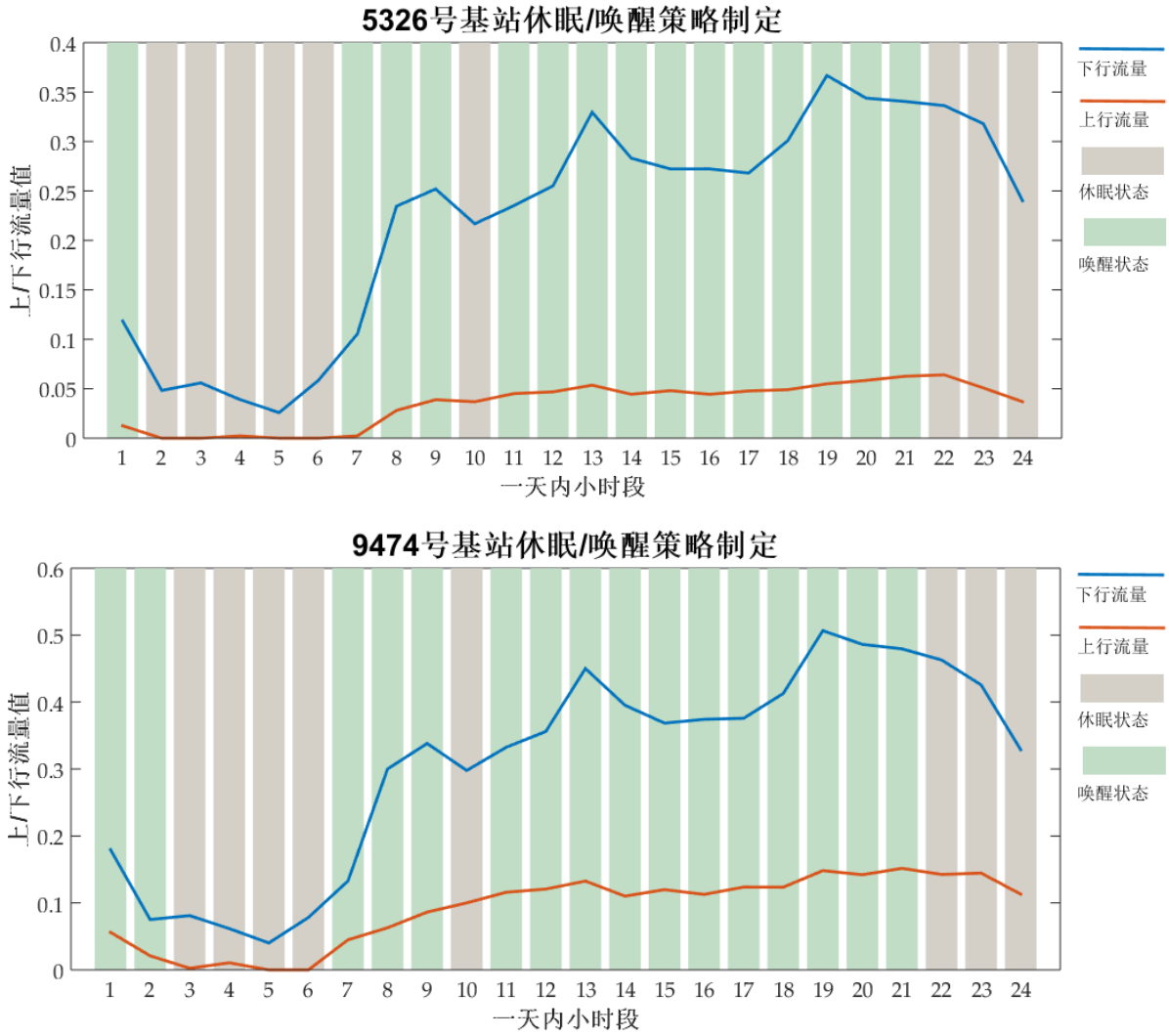


图 11. 5326 号和 9474 号基站休眠/唤醒策略制定

五、总结与讨论

本研究中总共给了 13 万+基站的数据，问题 1 主要针对这些基站进行分类并阐述相应类别的特点，问题 2 针对优化基站的运行设置合适的阈值和制定优越的策略。

其一，本文主要实现了基站的分类和相应类别特点的阐述。首先在数据层最需要进

行的是数据得预处理和特征数据处理，前者主要包括基站提取、时间戳处理、数据排序及清洗和时间段划分及统计，最终筛选出 118756 条可用数据；后者提取上下行流量高峰时段（Top3）、日均流量（GB）、日标准差、日峰值（GB）、连续变化绝对值之和以及上下行流量均值比作为小区基站历史流量数据的时序特征，共计 30 维特征。归一化后，输入到模型层，利用改进后的模糊 C 均值（FCM）非监督分类方法进行分类，其改进点在于首先利用类间距离和类内距离确定最佳分类数目为 12。而后，通过分类结果，计算各类别基站下数据的众数或者中位数用以 12 类基站特点的分析。

其二，本文主要针对优化基站的运行设置合适的阈值和制定优越的策略。以“基站休眠阈值的设置—基于长短期记忆网络（LSTM）的短期预测模型—基站休眠策略的制定”为研究主线，构建了基于流量负载预测的大规模移动通信基站休眠方法框架。其中，通过归一化负载限制值 β_1，β_2，β_3，将网络负载水平划分为四个等级，及超低负载水平、低负载水平、中负载水平和高负载水平。提前设置相应的休眠比例参数（z_1、z_2、z_3、z_4），根据短期预测结果，确定基站在未来各小时段内应采取休眠或者唤醒的状态，达到优化基站运行能耗和效率的目的。

针对现有研究进展，我们认为仍有以下两个问题可以值得进一步研究。1）基站网络中移动数据流量的季节性预测方法研究和 2）结合基站空间位置开展时空流量建模方法研究。

（注：相关程序和计算结果见附件压缩包）

六、参考文献

[1] 何勇，李艳婷，基于向量自回归模型的移动通信基站流量预测，工业工程与管理，22(4)：79-84，2017.

[2] 胡铮，袁浩，朱新宁，倪万里，面向 5G 需求的人群流量预测模型研究，通信学报，40(2)：1-10，2019.

[3] 陈沫，陈奎林，刘光毅，崔春风，新型无线接入网络架构研究，电信科学，27(1)：76-82，2011.

[4] Oh, E., Krishnamachari, B., Liu, X., Niu, Z. Toward dynamic energy-efficient operation of cellular network infrastructure, IEEE Communications Magazine, 49(6): 56-61, 2011.

[5] 刘濛，涂山山，肖创柏，林强强，一种基于模糊聚类的物理小区识别分配方案，现代电子技术，17，2019.

[6] Goonewardena, M., Akbari, H., Ajib, W., & Elbiaze, H. (2014, December). On minimum-collisions assignment in heterogeneous self-organizing networks. In 2014

IEEE Global Communications Conference (pp. 4665-4670). IEEE.

[7] Ning, L., Wang, Z., Guo, Q., & Zhang, H. Fuzzy layered physical cell identities assignment in heterogeneous and small cell networks. Electronics Letters, 52(10), 879-881, 2016.

[8] Wang, J., Shahidehpour, M., Li, Z., & Botterud, A. Strategic generation capacity expansion planning with incomplete information, IEEE Transactions on Power Systems, 24(2): 1002-1010, 2009.

[9] 任嘉鹏. (2020). 基于机器学习的流量预测及基站休眠方法研究 (Master's thesis, 吉林大学).

[10] 朱禹涛. (2016). 基于基站休眠的蜂窝网络能量效率优化技术研究 (Doctoral dissertation, 北京邮电大学).

[11] 李达. (2018). 5G 密集异构网络下的基站休眠技术研究 (Master's thesis, 北京邮电大学).

[12] 刘璐，吴成茂，基于类内类间距离的模糊 C-均值聚类分割算法，计算机工程与设计，37(6)：1626-1631，2016.

[13] Hochreiter, S., & Schmidhuber, J., Long short-term memory, Neural computation, 9(8): 1735-1780, 1997.