

DOI:10.20033/j.1003-7241.(2026)10-0078-06

基于 SOM-K 的 LTE 无线性能仿真研究

李 炜¹, 王少轩²

(1. 陕西国防工业职业技术学院 人工智能学院, 西安 710300; 2. 加泰罗尼亚理工大学, 西班牙 巴塞罗那 08034)

摘 要: 移动网络会产生海量的运营数据, 这些数据可以反映出包括多种和不同的网络性能测量和网络指标。如何通过这些数据分析预测无线网络性能成为了研究的热点。随着人工智能和大数据技术的发展, 对于海量数据的分析以及在其他领域的成功应用, 为本次在电信领域有效地监测小区性能和无线网络资源分配管理和处理的研究提供了可能。本次研究主要目的是通过人工智能机器学习算法对海量无线侧的运营数据进行数据挖掘分析, 形成适应于移动网络无线侧的人工智能聚类模型, 实现对移动通信小区的网络性能预测, 改善网络性能, 从而减小运营商网络优化开支。在此背景下, 分析了自组织映射 (self-organizing mapping, SOM) 和 K-means 算法的特点, 提出了一种融合 SOM 和 K-means 的人工智能聚类算法, 该方法用 SOM 算法训练输入数据, 将每个节点的权值向量作为 K-means 算法的中心向量以进行二次聚合, 从而更好地模拟移动网络性能。经实验表明和单独使用 K-means 方法相比, 该方法聚类效果更佳, 可以更好地表述移动网络性能。

关键词: 数据挖掘; 人工智能; 无线接入网; 自组织映射; K 均值聚类算法; 性能预测; SOM-K

中图分类号: TP391.9; TP18; TP311.13

文献标志码: A

文章编号: 1003-7241(2026)10-0078-06

Research on LTE wireless performance simulation based on SOM-K

Li Wei¹, Wang Shaoxuan²

(1. Dept. of Artificial Intelligence, Shaanxi institute of technology, Xi'an 710300, China;

2. Polytechnic University of Catalonia, Barcelona 08034, Spain)

Abstract: Mobile networks generate a vast volume of operational data, which encapsulates a wide array of network performance metrics and indicators. Predicting wireless network performance through the analysis of such data has emerged as a prominent research focus. The advancements in artificial intelligence and big data technologies, along with their successful applications in large-scale data analysis across diverse domains, have paved the way for effective research on cell performance monitoring and wireless network resource allocation management within the telecommunications sector. The primary objective of this study is to conduct data mining and analysis on massive wireless-side operational data using AI and machine learning algorithms, thereby developing an AI clustering model tailored to the wireless side of mobile networks. This model aims to enable accurate prediction of mobile communication cell performance, optimize network efficiency, and subsequently reduce network optimization expenditures for operators. Against this backdrop, the characteristics of the Self-Organizing Map (SOM) and K-means algorithms are analyzed, and a hybrid AI clustering algorithm integrating SOM and K-means is proposed. This method first trains the input data using the SOM algorithm, then leverages the weight vectors of each SOM node as initial centroid vectors for the K-means algorithm to perform secondary clustering, thereby enhancing the simulation of mobile network performance. Experimental results demonstrate that, compared with the standalone K-means algorithm, the proposed method achieves superior clustering performance and provides a more accurate representation of mobile network performance.

Keywords: data mining; AI; wireless access network; SOM; K-means clustering algorithm; performance prediction; SOM-K

随着大数据技术的发展, 大数据分析在电信运营商各级业务、服务和网络管理中的应用越来越广泛, 例如如何基于网络数据分析流量路由和服务质量参数选择, 如何在客户关注领域实现移动用户之间的数据处理是研究的热点问题^[1]。尤其是在移动通信网络无线则会源源不断地产生大量与业务相关的数据。因此, 数据挖掘和分析的兴起使电信运营商能够有效地监测移动通信小区性能和无线网络资源分配管理^[2]。聚类是一种简单而有效的人工

智能算法用来分析数据。因此, 引起了科研工作人员的广泛关注。它也是在许多领域(如工业、农业和经济系统等^[3])处理大型数据集的常规的标准方法之一。

聚类是一种用于将数据分组到具有相似特征的集合上的方法, 它可以用于观察数据的数据模式形成良好的集群。标记过的输入数据更容易实现聚类, 其结果可以作为一种惩罚/激励机制的判定, 以促进数据的有效聚类^[4]。在算法中聚类的 K-means 和基于样例(SOM)得到了学者的广泛研

收稿日期: 2024-09-10; 录用日期: 2024-10-11

基金项目: 陕西国防工业职业技术学院校本项目(Gfy25-09); 陕西国防工业职业技术学院项目(807/2112343)

作者简介: 李 炜(1990—), 男, 副教授, 硕士, 研究方向为人工智能、移动通信技术。

引用本文: 李炜, 王少轩. 基于 SOM-K 的 LTE 无线性能仿真研究[J]. 自动化技术与应用, 2026, 45(10): 78-83. (Li Wei, Wang Shaoxuan. Research on LTE wireless performance simulation based on SOM-K[J]. Techniques of Automation and Applications, 2026, 45(10): 78-83.)

究关注,在移动通信领域的监测也被广泛应用。例如,Andrades 等^[5]提出了一种 SOM-ward 聚类方法, Liu 等^[6]提出了一种具有低维标签数据集的 SOM+ *K-medoids* 方法用于 LTE 网络异常检测。Swarnalaxmi 等^[7]使用 *K-means* 聚类来检测来自 LTE 核心网络的标签数据集的任何异常行为,并通过添加其他参数来提高网络性能。本文此次研究与参考文献^[5]的工作非常相似,但文献^[5]基于低维数据集(即只使用二维数据集)提取了 3G 网络的行为模式,本文在无线接入网使用更复杂和更高维的数据集(即 29 维数据集)来分析 LTE 单元的性能。

1 SOM 和 *K-means*

SOM 本质是基于样例的方法,同时属于无监督竞争学习的神经网络模型。该方法的核心作用可实现数据的聚类、降维、可视化等,是一个简单、高度可视化的自动数据分析方法,广泛应用于异常检测、模式识别、数据挖掘领域中的数据聚类和分析^[8]。

SOM 聚类不同于其他神经网络,因为它采用的是竞争学习而不是纠错学习,而且在某种意义上,它使用邻域函数来保持输入空间的拓扑性质。SOM 算法优点是原理简单计算开销相对深度学习较低。保持原有数据的拓扑关系,易于实现数据的压缩,可将数据从高维空间映射到低维空间。

1.1 *K-means*

K-means 聚类是最经典、使用最为广泛的一种聚类算法。首先把数据集划分成预先制定的 *K* 个簇,通过不同类别的迭代^[9],让同一个簇内的样本距离较为相近。基于样本的相似性,使内部集群的样本差异较小,而不同集群的样本差异大,这是一个动态聚类算法适合中小型数据聚类。然而,*K-means* 存在固有的缺点:1) *K*-均值算法要求必须提前给定 *K* 值,因无法自动判断聚类个数,*K* 的值往往难以确定;2) 依赖初始质心,容易陷入局部最优;3) 无法处理分类特征、故,本文结合两个算法的优缺点提出了一种结合 SOM 和 *K-means* 的两段式聚类方法。

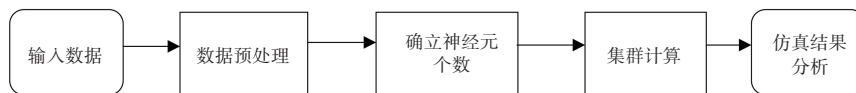


图 1 聚类流程

Fig. 1 Clustering process

在实施 SOM 模型训练流程前,归一化是输入数据预处理的必要步骤。各类特征原始取值区间存在明显区分,不加处理直接导入模型会诱发尺度偏差问题。标准化公式实现如下:

$$Z_{ij} = \frac{x_{ij} - \min x_j}{\max x_j - \min x_j} \quad (4)$$

式中, x_{ij} 为数据项 i 中特征 j 的原始值, $\min x_j$ 、 $\max x_j$ 分别为特征 j 在整个数据项中的最小数据值和最大数据值, Z_{ij} 为归一化的值,归一化的结果在 $[0,1]$ 之内。

1.2 SOM

SOM 网络结构由输入层和竞争输出层组成。SOM 的工作原理是将从输入层的神经元中输出的数据投射到竞争输出层的神经元中^[10]。在此过程中,当一个输入样本被输入到 SOM 模型时,计算其与所有权值向量的欧氏距离,并将最佳匹配单元(best matching units, BMU)和 SOM 网格中靠近它的其他神经元的权值调整到输入向量^[11-13]。SOM 的基本训练过程如下。

1) Y_i 随机选择输入数据集: $Y(t) = Y_i$ 的一个新样本, i 在 1 和 I 之间随机选择。

2) 寻找 $Y(t)$ 的最佳匹配单元(best matching unit, BMU):在竞争层中搜索一个神经元 $q \in 1, \dots, N$ 。通过计算输入数据与每个神经元权重的欧氏距离获得 BMU,距离最小的为获胜单元。

$$\min_p \{ \|U_p - Y(t)\| \} = \|U_q - Y(t)\| \quad (1)$$

式中 $\|\cdot\|$ 是距离测量。

3) 更新权重,对 BMU 及其邻域内的神经元参数的权重进行修复,减小相对距离。

$$U_p(t+1) = U_p(t) + \beta(t) \times g_{ci}(t) [Y(t) - U_p(t)] \quad (2)$$

$U_p(t+1)$ 是更新的权重向量, $\beta(t)$ 为学习速率,其取值范围一般是从 0 到 1,随迭代逐步减小,保证后期稳定收敛。

$$g_{ci}(t) = \exp - \frac{\|r_i - r_q\|^2}{2\sigma(t)^2} \quad (3)$$

$g_{ci}(t)$ 是邻域函数,通常表示为一个高斯函数,以 BMU 为中心。其中, r_i 是 BMU 坐标, r_q 是其他的神经元坐标,当 t 趋近于 ∞ 时, $g_{ci}(t)$ 必须接近于 0 才能使算法收敛。 $\sigma(t)$ 为邻域半径,随训练轮次 t 线性衰减在机器学习过程中,学习速率和邻域函数的宽度会降低,通常呈线性方式。

4) 重复以上步骤,直到达到预设的迭代次数,训练结束。

1.3 SOM-K 聚类

SOM 算法用作聚类的流程如图 1 所示。本次,输入数据为 29 维,数据项中最大数量是为 $N = 85\ 977$ 。

将输入数据完成归一化后就需要确定 SOM 模型竞争层中合适的神经元个数 N 。可以通过计算平均量化误差(quantization error, QE)和拓扑误差(topographic error, TE)来确立决定神经元数量 N ,这是评估 SOM 模型质量的主要测量方法^[16-19]。量化误差(QE)可用于统计全部的输入样本向量和其对应的最优匹配神经元之间距离的均值。拓扑误差(TE)则用来统计次优获胜神经元和最优匹配单元不存在邻接关系的样本占比^[20-22]。从评价标准来看, QE、TE 两项指标数值越低,代表 SOM 网络的拓扑拟合效

果与整体聚类表现越优异。然而,在决定神经元数量时, QE 和 TE 之间的相关效应必须考虑到。

$$QE = \frac{1}{R} \sum_{t=1}^R \|Y(t) - W_q(t)\| \quad (5)$$

$$TE = \frac{1}{R} \sum_{t=1}^R d(Y(t)) \quad (6)$$

式中, $Y(t)$ 为第 t 次迭代的输入数据项; $W_q(t)$ 为样本 $Y(t)$ 的 BMU 的权值向量; $d(Y(t)) = 1$ 时, 如果 $Y(t)$ 第 1 次的 BMU 和第 2 次的 BMU 不相邻, 则认为 R 为网络收敛前的迭代次数且 $d(Y(t)) = 0$ 。

基于以上所述, 在这一步中, 本文使用 MatlabSOM 工具箱 2.0 训练 SOM 模型, 其中不同数量的神经元从 4 个 (即 2×2 网格) 到 100 个神经元 (即 10×10 网格)。QE 和 TE 的值对于每个配置, 使用软件包计算。在所有训练过的模型中, 将选择 QE 和 TE 最低的模型。在 TE/QE 方面最优的 SOM 模型的大小表示为 P_{opt} 。

接下来, 借助 K -means 聚类算法的优势, 对自组织映射网络 (SOM) 输出的最优权向量集 V_{opt} 实施二次聚类, 最终生成数量更少的特征向量集合 Q , 且满足 $Q \leq V_{opt}$, 以此提升改进模型整体的聚类划分效果。这组数量为 Q' 的向量即为各类簇的聚类质心。算法通过戴维森 - 伯丁指数筛选最优聚类数量 T , 该评价指标以簇内样本间距与簇间中心间距的比值作为衡量依据, 数值越低代表聚类划分效果越理想。融合 SOM 与 K -means 的组合模型属于双层聚类模型, 既可以发挥 SOM 的数据自组织映射优势, 又能发挥 K -means 运算速度快的特点; 同时 SOM 预处理后输出的权向量, 可为 K -means 提供收敛区间更稳定、取值范围更集中的初始聚类中心点。

Q 簇的戴维森 - 伯丁指数由以下表达式定义^[23-27]。

$$S_{DB}(Q) = \frac{1}{Q} \sum_{m=1}^U \max_{l \neq m} \left\{ \frac{S_c(Q_m) + S_c(Q_l)}{d_{ce}(Q_m, Q_l)} \right\} \quad (7)$$

其中:

1) Q_m 和 Q_l 表示 Q 簇集合中的第 m 个和第 l 个簇。

2) $S_c(Q_m) = \frac{\sum_i \|Y_i - C_{Q_m}\|}{N_{Q_m}}$ 是质心团簇内距离的计算。

$S_c(Q_m)$ 是从簇的样本到簇质心的所有成对距离的平均值。 X_i 是分配给 Q_m 聚类的 n 维特征向量。 C_{Q_m} 是簇 Q_m 的质心。

3) $d_{ce}(Q_m, Q_l)$ 是两个簇 Q_m 和 Q_l 之间的簇间距。DBI 可以通过对 Q 一些数据计算得到, 取最小化 $v_{DB}(Q)$ 的值作为最优值, 记为 Q^* 。

4) 最后, 通过比较各聚类不同特征的分布情况分析不同聚类的特征。

2 实验与分析

本文试图根据在整个测量收集期间观察到的小区长期行为 (长期行为是在整个 15 d 内的小区模型的表现) 来获得小区模型。首先计算所有单元的整个测量周期的

平均值, 将输入数据选择到 63 行向量 (每个单元一个), 每个单元的 29 列向量 (每个特征一个)。本文将输入数据表示为一个 63×29 的矩阵。在仿真工作中, 本文也尝试了多次组迭代 (例如, 从 50 到 1 000), 发现当迭代时间为 350 时, 仿真的精度是稳定的, 可以达到 0.92)。

为了确定拓扑的大小, QE 和 TE 的值计算了从 2×2 神经元到 10×10 神经元的配置。所得到的值如表 1 所示, 可以看到, 当配置达到 6×6 时, QE 和 TE 得到最小值 (QE=0.296 和 TE=0)。

表 1 QE 和 TE 值

Tab. 1 QE and TE value

拓扑大小	QE	TE
2×2	1.66	0
3×3	0.79	0
4×4	0.95	0
5×5	0.76	0
6×6	0.296	0
7×7	0.46	0
8×8	0.43	0
9×9	0.4	0
10×10	0.41	0.5

与每个神经元相关联的权值向量可视作为一个输入向量簇的中心。此外, 拓扑中彼此相邻的神经元在输入层中相互靠近, 因此, 可以在二维的网络拓扑中可视化一个高维的输入层数据。SOM 的一个可视化工具是权重距离矩阵 (也称为 u 型矩阵), 如图 2 所示。蓝色的六边形代表神经元。红线连接着相邻的神经元。包含红线的区域中的颜色表示神经元之间的距离。 u 型矩阵其元素表示每个映射小区与其相邻小区的平均距离, 因此这个距离矩阵的相对最大值表示聚类边界, 而较低的值表示聚类本身。

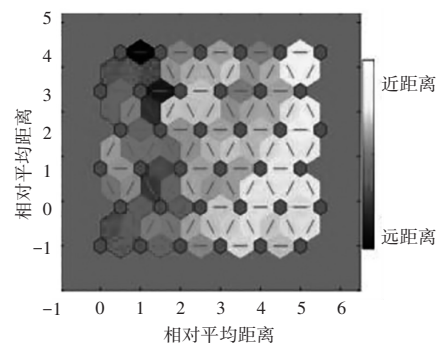


图 2 邻重距离

Fig. 2 Neighbor weight distance

SOM 样本命中结果如图 3 所示。每个神经元显示了其分类的输入向量的数量。每个神经元的向量的相对数量由彩色斑块的大小表示, SOM 的默认拓扑结构是六边形的。这个图显示了神经元在拓扑结构中的位置, 并显示了有多少训练数据与每个神经元相关联。拓扑结构是一个 6×6 的网格, 所以有 36 个神经元。SOM 的初始簇数为 35, 与任何神经元相关的最大命中数为 4。

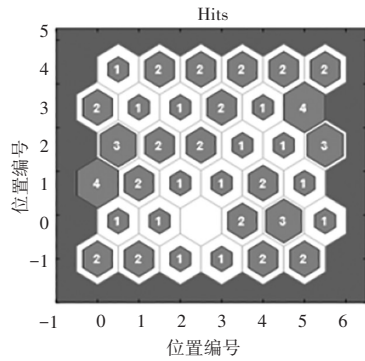


图 3 SOM 样本命中结果

Fig. 3 SOM hits of each neuron

为了方便将输入划分为不同的聚类结果,本次采用 SOM 算法的权值平面可视化拓扑结构;从 9 个不同类别

中选择的 9 个不同特征的权值如图 4 所示。它们也代表了不同特征之间的相关性。与此同时,它们是连接每个输入到每个神经元的权重的可视化(深色的权重越高,越浅的权重越低)。如果两个输入的连接模式(例如,颜色变化和分布相似)非常相似,则可以假设输入高度相关。例如,图 4(a)、4(b)、4(c)、4(d)的颜色变化和分布几乎相似,因此,可以确定平均 UEDL(下行)和数据量 DL、最大吞吐量 DL 和 PRB 使用 DL 呈正相关。图 4(f)、4(i)的颜色分布几乎相似,这说明延迟 DL 和 RRC 设置失败率有很强的相关性。然而,图 4(e)与其他 8 个特征有不同的颜色变化和分布,因此,可确定 IntereNBHOS1 失败率与其他 8 个特征之间的相关性较弱。另一方面,图 4(f)、4(g),图 4(g)、4(i)在每个神经元中的颜色分布都相反,这意味着这 2 组的特征分别具有负相关特征。

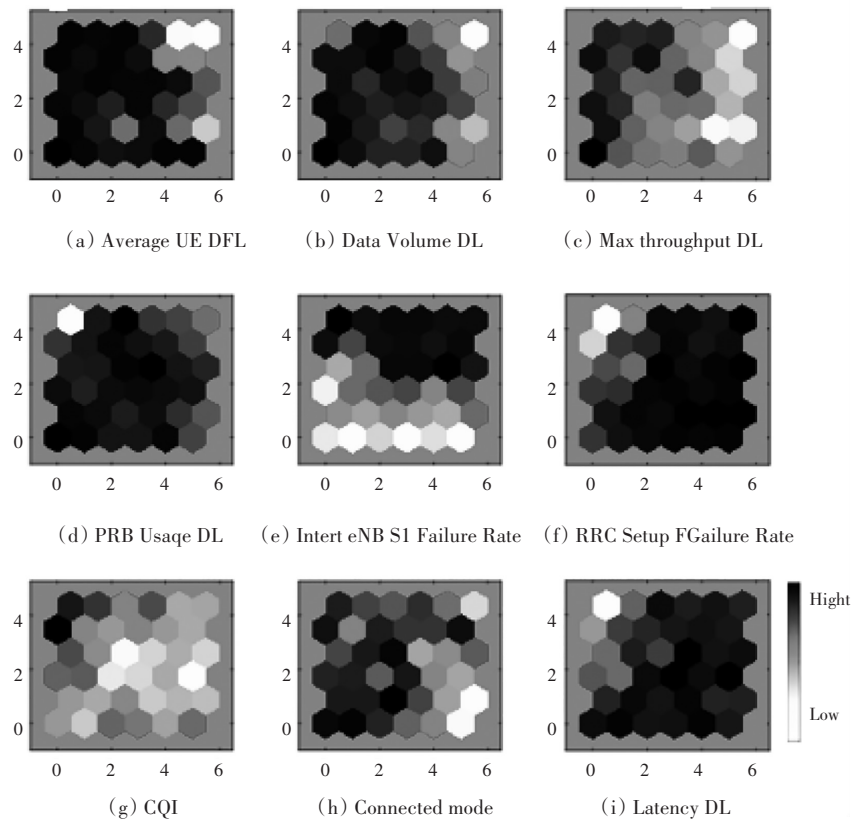


图 4 不同特征的权重

Fig. 4 Weight planes of different features

对于特定的聚类,可以从图 5 中找到。表示 DBI 的最小值为 2.1,对应的簇覆盖小区数为 5。如图 5 所示,本文使用不同的颜色来代表不同的簇。

经过以上实验验证,此次采用的 SOM-K 二段式聚类算法能够较好地预测 LTE 小区性能。最后,将 K-means 和 SOM-K 算法进行比较,并将它们的聚类数设置为 5 个聚类,如表 2 所示。

表 2 展示了 K-means 和 SOM-K 不同算法的簇内和簇间质心距离。数据表明 K-means 算法的簇内距离相对较大,簇间距离相对较小;依据“簇内紧凑、簇间分离”准则^[11],高维数据下 SOM-K 聚类性能优于 K-means。

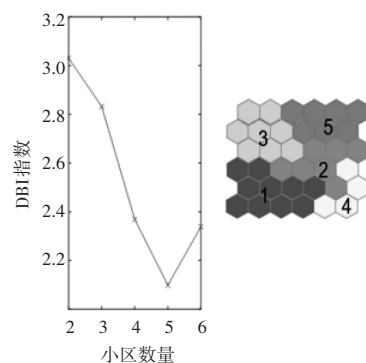


图 5 DBI 和响应集群的最小值

Fig. 5 The minimum value of DBI and responding cluster

表 2 聚类簇质心距

Tab. 2 Distance comparison between *K*-means and SOM-K

集群编号	簇内质心距离 (<i>K</i> -means)	簇内质心距离 (SOM-K)	簇间质心距离 (<i>K</i> -means)	簇间质心距离 (SOM-K)
1	1.05	0.405	1.01	1.08
2	0.63	0.577	0.59	0.88
3	0.68	0.655	0.53	0.96
4	1	0.505	0.63	0.80
5	0.52	0.470	0.71	0.85

将表 3 和表 4 进行比较可以发现,SOM-K 中的集群#5 与 *K*-means 中的集群#1 和集群#4 具有相似的特征。这都是整个 5 个集群中性能最好的。同样的情况也分别发生在 SOM-K 中的集群#1 和#2 以及 *K*-means 中的集群#3 和#5 上。SOM-K 中的集群#2 和单个 *K*-means 中的集群#5 具有相似的性能,这是所有最差的集群。SOM-K 中的集群#1 和 *K*-means 中的集群#3 都是中等性能的集群。这两

种算法之间唯一的区别是 SOM-K 中的集群#4 和 *K*-means 中的集群#2,例如,SOM-K 中的集群#4 具有较低的 RRC 设置失败率和延迟,而 *K*-means 中的集群#2 具有较高的 RRC 设置失败率和延迟。簇 2 中的小区数量大于簇 4。综上所述,这也可以说明 *K*-means 聚类在高维数据集上的性能不佳,首先应用降维方法(即 SOM)对数据进行降维,然后使用 *K*-means 会有更好的聚类效果。

表 3 SOM-K 算法不同簇中各小区性能

Tab. 3 Number of cells in different clusters by using SOM-K

集群编号	单元数	特征
1	16(60,62,56,21,38,35,37,57,18,5,6,25,44,43,48,49)	中:UEs、数据量、吞吐量、RRC 设置失败率、PRB 使用量、CQI、连接模式和延迟。 低:eNB、HO、S1 之间故障率。
2	12(61,54,63,29,47,51,33,14,41,42,17,1)	高:RRC 设置失败率和延迟。 低:UES、数据量、吞吐量、PRB 使用率、CQI 连接模式,eNB 和 HO、S1 之间故障率。
3	14(32,34,46,9,15,7,8,19,58,59,10,50,4,11)	等:UEs、CQI、数据卷、吞吐量、RRC 设置失败率、连接模式和 PRB 使用情况。 低:延迟和 eNB、HO、S1 间故障率。
4	6(20,55,3,2,23,24)	等:已连接模式和 CQI。 低:UEs、数据量、吞吐量、eNB 间 HOS1/RRC 设置失败率、PRB 使用情况、延迟。
5	15(39,16,30,31,36,45,12,27,26,22,28,13,52,53,40)	高:UEs、数据量、吞吐量、PRB 使用量、CQI 和连接模式。 中:延迟。低:RRC 设置失败率。

表 4 *K*-means 算法不同簇中的小区特征

Tab. 4 Different clusters by using *K*-means

集群编号	单元数	特征
1	14	高:UEs、数据量、吞吐量、PRB 使用量、CQI 和连接模式。 中:延迟;低:RRC 设置失败率。
2	20	高:RRC 设置失败率和延迟。中:PRB 的使用量和 CQI。 低:UEs、数据量、吞吐量、eNB 和 HO、S1 间故障率和连接模式。
3	12	中:UEs、数据卷、吞吐量、RRC 设置失败率、PRB 使用情况、CQI、连接模式和延迟。 低:eNBHOS1 故障率。
4	9	高:UEs、数据量、吞吐量、CQI。 中:PRB 的使用量。 低:RRC 设置失败率和延迟。
5	8	高:RRC 设置失败率和延迟。 中:PRB 的使用量和 CQI。 低:UEs、数据量、吞吐量。

3 结论

基于SOM-K预测LTE无线网络性能的这种方法尚未得到推广使用。本次研究可以表明SOM-K算法可以表征LTE小区性能状况。同时,基于SOM-K的分析方法的优点是在分析中采用了2步测量方法来提高聚类结果。此外,该方法在网络性能的聚类效果明显优于单独使用K-means算法。总之,在未来对LTE网络进行优化时可以使用SOM-K方法来预测网络性能,同时帮组运营商智能地了解和提前预判网络性能。

在未来,B5G/6G网络的运行将主要由服务驱动,B5G/6G网络的趋势将进一步增加数据的使用。值得注意的是,基于异常网络行为分析的SOM-K检测已长期成功地应用于GSM数据中。因此,探索该算法在B5G/6G领域的无线网络性能预测和网络性能的管理(数智化、自动化网络节能降耗)将成为一个新的研究方向。

参考文献

- [1] 邓扬鑫,赵凤,朱晓荣.基于大数据挖掘的LTE网络弱覆盖分析及覆盖优化研究[J].南京邮电大学学报,2021,41(2):25-33.
- [2] 邢旭东.基于大数据的蜂窝网络流量变化趋势预测研究[D].北京:北京邮电大学,2021.
- [3] 汪海涛,余松森.分布式SOM结合K-均值聚类的软件定义网络泛洪攻击检测方法[J].计算机应用研究,2019,36(11):3423-3427.
- [4] 钟超,刘漫,丹贺帆.基于改进密度峰值算法的轨迹聚类[J].计算机工程与设计,2024,45(1):130-138.
- [5] Andrades A G, Muñoz P, Serrano I, et al. Automatic root cause analysis for LTE networks based on unsupervised techniques[J]. IEEE, 2016 (65)4:2369-2386.
- [6] Liu X, Chuai G, Gao W D, et al. KQIs-driven QoE anomaly detection and root cause analysis in cellular networks[C]//2019 IEEE Globecom Workshops. Waikoloa, HI: IEEE, 2019:1-6.
- [7] Swarnalaxmi S, Elakkiya I, Thilagavathi M, et al. User activity analysis driven anomaly detection in cellular network[C]//2018 Tenth International Conference on Advanced Computing (ICoAC). Chennai, India: IEEE, 2018:159-163.
- [8] 李炜,王少轩. SOM在LTE小区性能分析中的应用[J].微处理机, 2022,43(4):38-41.
- [9] 杨芳,湛燕,田学东,等.使用遗传算法实现K-means聚类算法的K值选择[J].微机发展,2003,13(1):25-26.
- [10] 欧逸,哲术茜.基于SOM和K均值聚类的谐振接地系统故障选线及区段定位方法[J].电气技术,2023,24(10):23-30.
- [11] Liu Y C, Li Z M, Xiong H, et al. Understanding of internal clustering validation measures[C]//2010 IEEE International Conference on Data Mining. Sydney, Australia: IEEE, 2010:911-916.
- [12] 周欢,李广明,张高煜. SOM+K-means两阶段聚类算法及其应用[J].现代电子技术,2010,33(16):113-116.
- [13] 董俐,君岳恒.基于大数据挖掘的用电异常特征自动化分析方法[J].自动化技术与应用,2025,44(10):125-128.
- [14] 刘才华,苏鑫磊,刘鹏源,等.基于主成分聚类分析的智能配电网信息自动集成方法[J].自动化技术与应用,2026,45(2):128-132.
- [15] 刘乐,张显彰,钱堃,等.基于烟花算法优化K均值聚类的家电负荷需求响应潜力评估[J].自动化技术与应用,2025,44(6):185-188.
- [16] 史国斌,张忠林.基于改进层次方法的无参K-means算法[J].计算机真,2022,39(8):340-346,426.
- [17] 郑星航,张京伦,蔡晴,等.基于多源数据融合的电力企业智慧客服系统[J].自动化技术与应用,2023,42(5):154-157.
- [18] 苏立伟,刘振华,苏华权,等.基于数据负荷序列聚类的配电网电力营销实时信息系统测试[J].电网与清洁能源,2021,37(12):64-69,80.
- [19] 颜子寒,张正军,王雅萍,等.基于加权马氏距离的改进深度嵌入聚类算法[J].计算机应用,2019,39(增刊2):122-126.
- [20] 吕维,钱宇华,王婕婷,等.结合力导向图分布算法的特征加权深度嵌入聚类[J].小型微型计算机系统,2024,45(6):1318-1324.
- [21] 李秋云,刘燕武.一种服务于K-means的初始中心选取方法[J].电子技术应用,2023,49(3):134-138.
- [22] 罗博伟,罗万红等.融合SOM神经网络与K-means聚类算法的用户信用画像研究[J].铁路计算机应用,2024,33(7):14-19.
- [23] 胡津铭,张艳.基于改进SOM神经网络的移动警务终端流量监测方法研究[J].信息安全,2021(增刊1):109-112.
- [24] 熊健,覃仁超,何梦乙,等.改进随机森林在Android恶意软件检测中的应用[J].计算机工程与应用,2021,51(3):10-19.
- [25] 丁江桥,文屹,吕黔苏,等.基于时间序列和神经网络的电力设备状态异常检测方法[J].电测与仪表,2024,61(2):185-190.
- [26] 宋金伟,李俊妮,宣东海,等.基于改进K-medoids聚类和碳约束的变压器状态异常大数据诊断方法研究[J].电测与仪表,2026,63(5):20-29.
- [27] 杨威超,郭渊博,钟雅,等.基于设备型号分类和BP神经网络的物联网流量异常检测[J].信息安全,2019,19(12):53-63.