



基于聚类分析和深度学习的多频多模网络负载均衡优化

邱亚星¹, 王希栋¹, 边森¹, 岳磊²

(1. 中国移动通信有限公司研究院, 北京 100053;
2. 中国移动通信集团广西有限公司, 广西 南宁 530022)

摘要: 负载均衡问题是 LTE 多频多模网络要解决的重大问题。多频多模网络结构复杂, 负载均衡涉及的参数达数百个, 仅依靠人工经验很难进行精细化配置。为解决多频多模网络的负载均衡问题, 解决现网运维的难点与痛点, 提出一种基于机器学习的多频多模网络负载均衡方案。首先选取关键指标对网络场景进行划分, 然后利用机器学习技术挖掘出不同场景下的最佳参数配置建议。经验证, 机器学习技术可以大大提高参数配置的质量和效率, 做到精细化参数配置。

关键词: 多频多模网络; 机器学习; 负荷优化

中图分类号: TN929.53

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2020159

Load balancing based on clustering analysis and deep learning for multi-frequency and multi-mode network

QIU Yaxing¹, WANG Xidong¹, BIAN Sen¹, YUE Lei²

1. China Mobile Research Institute, Beijing 100053, China
2. China Mobile Group Guangxi Co., Ltd., Nanning 530022, China

Abstract: Load balancing is a huge challenge for LTE multi-frequency and multi-mode network. Hundreds of parameters are involved in load balancing for the complex network structure. Therefore, it is difficult to perform precise and meticulous configuration only relying on human experience. In order to cope with the challenge, a load balancing scheme based on clustering analysis and deep learning was proposed. Firstly, the key indicators were selected to identify the network scenes, and then big data and deep learning technologies were used to mine the relationship between data. Finally, the optimum system parameters for different network scenes were found. It has been proved that machine learning technology can greatly improve the accuracy and the efficiency of parameter configuration.

Key words: multi-frequency and multi-mode network, machine learning, load optimization

1 引言

无限流量卡的推广导致 LTE 业务流量激增,

现有“F+D+E”的网络厚度不足以满足业务需求。“TDD+FDD(TF)”融合组网方式可以解决容量、深度覆盖、竞争三重网络挑战, 当

前网络 FDD 1 800 MHz 终端渗透率为 60%~70%，FDD 900 MHz 终端渗透率为 10%~20%，TF 深度融合可提升网络竞争力，应对业务流量激增的挑战。

负载均衡问题是 TF 深度融合多频多模网络所要解决的重大问题。针对多频多模、复杂的网络结构，通常一个扇区组配置一套参数，用来做负载均衡。目前的技术方案主要以网络优化工程师的经验为准绳，针对多频多模网络的具体频点组成，对资源配置、频点配置以及边界电平等相关参数进行配置。由于不同场景下的参数配置与网络质量的相关性十分复杂，仅依靠人工经验很难做到精细化的参数配置，并且人工经验没有依据大量数据，仅参考碎片化的小量数据进行经验总结，难以挖掘数据之间的隐藏关系以及利用这些关系提高网络运维的质量和效率，人工解决负载均衡问题的成本高且精确度低。

为了有效解决多频多模网络的负载均衡问题，本文提出一种基于聚类分析和深度学习的多频多模网络负载均衡方案，针对多频多模的特定网络场景，选取关键指标对场景进行划分，然后利用深度学习技术挖掘不同场景下的配置参数与网络质量之间的隐含关系，并对不同场景进行个性化的深度学习模型训练，依据训练得到的模型给出相应的参数配置建议。经现网测试验证，机器学习技术可以大大提高参数配置的质量和效率，做到精细化负载均衡。

2 多频多模网络负载均衡优化

2.1 TF 融合组网

TF 融合的多频多模网络结构如图 1 所示，整体组网主要由 3 部分组成：容量层、覆盖层、广深覆盖层。容量层主要用于业务吸收，其中 D 频段覆盖能力较差，用于近点、热点话务吸收，FDD 1 800 MHz 承担热点均衡效果，与 TDD 2.6 GHz 的 D 频段共同承担容量层。F 频段覆盖次之，用于基

础覆盖以及小区边缘业务吸收。FDD 900 MHz 频段覆盖效果最好，优先级最低，用于城区深度覆盖或郊区广覆盖。

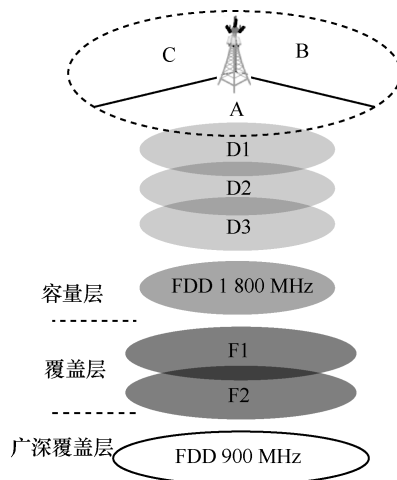


图 1 TF 融合的多频多模网络结构

TF 融合组网使得网络厚度增加，可快速、有效地提升网络能力，但日益复杂的网络结构却给负荷优化带来了严峻的挑战。不同频点之间的负荷均衡与重选边界主要通过切换、重选等相关参数控制，涉及参数可达数百个。仅依靠人工经验很难进行精细配置，参数配置与网络质量的相关性又十分复杂，不仅耗费大量的人力，而且很难达到精细化的效果。所以，传统的仅依靠网络优化工程师的经验进行负荷优化的方法不能适用于多频多模网络结构，同时，传统基于专家经验的优化方案也不具备现网大规模应用的可实施性。

2.2 基于聚类分析和深度学习的负载均衡优化

用户在多频多模网络中不同频点之间的切换、重选主要通过网络的相关参数控制，如何对相关数百个参数进行精确且有效的配置是影响负载均衡的关键因素。本文利用机器学习技术调整网络配置参数，以实现在容量层、覆盖层和广深覆盖层之间按小区承载能力均匀分配网络业务负荷。同时，依据小区用户分布特点划分小区类型，为不同类型小区生成个性化优化方案，且负载均衡方案可随网络业务发展变化实现自动化更新。基于

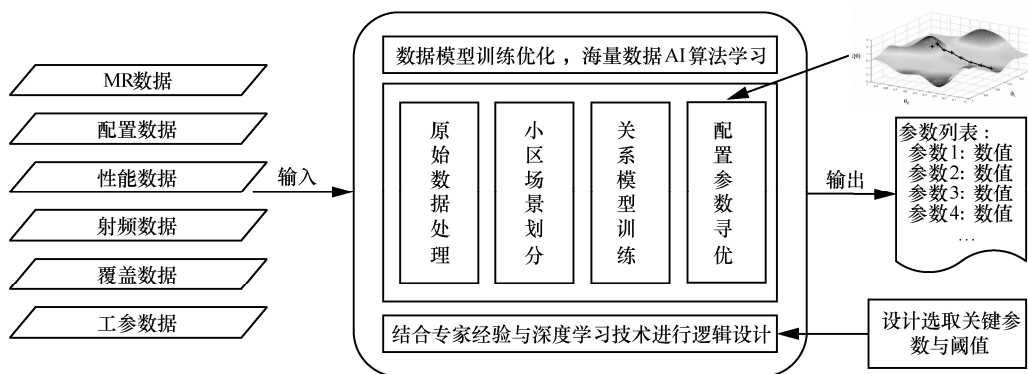


图2 基于机器学习的负载均衡优化流程

机器学习的负载均衡优化流程如图2所示。

(1) 原始数据处理

获得的数据为一定时间段内的小区级数据，包括公参数据、测量报告（measurement report, MR）数据、配置数据以及性能数据等。基于大数据平台的存储资源和计算资源完成数据的 ETL（extract-transform-load, 抽取、转换、加载）处理，为后续分析与建模提供可用的格式化数据。

(2) 小区场景划分

本部分主要实现针对小区用户分布特性进行场景划分，小区的 MR 数据可以较好地反映小区的用户分布情况，通过对 MR 数据的分析处理挖掘小区的潜在相似模式。MR 数据是按 RSRP（reference signal receiving power, 参考信号接收功率）强度分段统计的，按照 RSRP 强度、MR 数量绘制二维曲线，具体如图3所示。

从 MR 数据中提取用于小区场景划分的关键特征，提取每条 MR 数据的算术平均数、方差、峰度系数以及偏态系数。使用提取的这4个特征进行场景划分，场景自动划分采用 K-means 聚类算法，定义小区数量为 N ，且每个小区的 MR 数据已提取以上4个特征，则数据矩阵如下：

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ \dots & & & \\ x_{N1} & x_{N2} & x_{N3} & x_{N4} \end{pmatrix} \quad (1)$$

其中， x_{ij} 表示第 i 个扇区的第 j 个属性。

首先随机从数据集中选取 K 个点作为初始聚类中心，然后计算各个样本到聚类中心点的距离，把样本归到离它最近的那个聚类中心所在的类。采用欧氏距离计算数据对象间的距离：

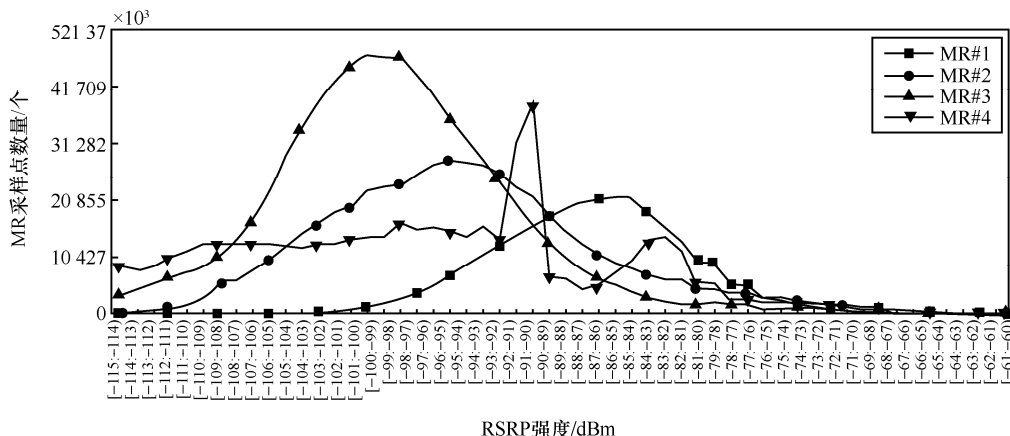


图3 MR 采样点数在 RSRP 强度区间的分布示例

$$\text{dist}(x_i, x_j) = \sqrt{\sum_{d=1}^4 (x_{i,d} - x_{j,d})^2} \quad (2)$$

更新类簇中心：对应类簇中所有数据对象的均值，即更新后该类簇的类簇中心。定义第 k 个类簇的类簇中心为 Center_k ，则类簇中心更新方式如下：

$$\text{Center}_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i \quad (3)$$

其中， C_k 表示第 k 个类簇， $|C_k|$ 表示第 k 个类簇中数据对象的个数，这里的求和是指类簇 C_k 中所有元素在每列属性上的和，因此 Center_k 也是一个含有 4 个属性的向量，表示为：

$$\text{Center}_k = (\text{Center}_{k,1}, \text{Center}_{k,2}, \dots, \text{Center}_{k,4}) \quad (4)$$

不断地迭代来重新划分类簇，并更新类簇中心，设定迭代次数 T ，第 T 次迭代后则终止迭代，此时所得类簇即最终聚类结果。

(3) 关系模型训练

基于 (2) 中所划分的场景，挖掘各个场景下的配置参数对性能指标的影响和配置数据与性能数据之间的隐含关系，调整无线参数，改善网络性能，提高网络 KPI (key performance indicator, 关键性能指标)。利用 DNN (deep neural network, 深度神经网络) 模型学习配置参数与性能指标参数的内在关系，得到训练稳定的神经网络模型。

定义配置参数 X 包括整个小区组的整套参数配置，包括其所包含的各个小区的资源配置参数、频点配置参数和边界电平配置参数等，包含 m 个频点、每个频点有 n 个参数的矩阵可表示为：

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & & & \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \quad (5)$$

其中， x_{mn} 表示第 m 个频点的第 n 个参数。如 F1D1D2FDD 900 MHz 的组网结构包含 4 个频点，每个频点如果有 12 个参数配置数据，则 X 矩阵大

小为 (4,12)。

性能指标参数 Y 为小区组负载均衡的判定指标数据，可包括：小区组所包含每个频点的 RRC (radio resource control, 无线资源控制) 最大连接用户数、上行 PRB (physical resource block, 物理资源块) 利用率和下行 PRB 利用率等参数，包括 m 个频点，每个频点有 k 个参数，矩阵大小为 (m,k) ，可表示为：

$$Y = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1k} \\ y_{21} & y_{22} & \dots & y_{2k} \\ \dots & & & \\ y_{m1} & y_{m2} & \dots & y_{mk} \end{bmatrix} \quad (6)$$

其中， y_{mk} 表示第 m 个频点的第 k 个负载指标数据。如 F1D1D2FDD 900 MHz 的组网结构包含 4 个频点，每个频点有 3 个指标数据，则 Y 矩阵大小为 (4,3)。

研究性能指标与配置参数的关系，将此问题建模为多变量回归问题。使用全连接神经网络算法模型学习回归问题的内在关系，使用训练数据训练神经网络模型。输入为 X ，读入 X 中的数据，转换为列矩阵，作为神经网络的输入，同理转换 Y 矩阵为列矩阵，作为神经网络的输出。

回归问题的输出 Y 含有多个变量，因此依据此问题设计代价函数为：

$$\text{MSE}(y, y') = \frac{\sum_{i=1}^l (y_i - y'_i)^2}{l} \quad (7)$$

其中， y_i 是样本数据，而 y'_i 是神经网络计算的预测值。

将各个变量的 MSE (mean squared error, 均方误差) 的平均值作为损失函数，通过训练模型、调整权重和偏置使得网络的代价函数达到最小，由此得到训练稳定的神经网络模型。

(4) 配置参数寻优

使用优化搜索算法在得出的训练模型中搜索使网络性能参数达到最优的参数配置，采用爬山



算法进行局部最优解的搜索，将得出的参数配置部署到现网中并测试其性能。优化配置输出过程如图 4 所示，具体算法实现为：从当前的节点开始，和周围的邻居节点的值进行比较。如果当前节点是最大的，那么返回当前节点，作为最大值（即山峰最高点）；反之用最高的邻居节点替换当前节点，从而实现向山峰的高处攀爬的目的。如此循环直到达到迭代次数。在迭代次数内，得到的可能只是局部最优解，并非全局最优解。



图 4 优化配置输出过程

3 现网测试验证

为验证算法的有效性与可实施性，需采集现网的数据集对本文提出的方案进行模型训练与结果评估，然后依据训练生成的算法模型与评估结果对算法模型进行现网部署与测试验证。

3.1 现网数据采集与预处理

采集现网中的数据，主要包括：MR、影响负荷的配置数据、与负荷相关的 KPI 与 PM 数据等。测量报告数据主要来自 UE 和 eNode B 的物理层、RLC 层以及在无线资源管理过程。影响负荷的配置数据指主要用于系统中无线资源控制子层完成诸如小区选择/重选及切换等事件触发的参数。与负荷相关的 KPI 与 PM 数据主要包括网络考核负荷所用的关键指标，如 RRC 最大连接用户数、上行 PUSCH 利用率、下行 PDSCH 利用率等。

为了满足机器学习算法的多维数据需求，需从不同设备厂商以及不同网元采集数据，这样采集的数据格式差异大、数据相关性复杂，关联去重之后实际可用格式化数据减少。本文所选取的小区场景为 F1D1D2FDD 900 MHz，采集南宁市中存在负荷不均衡的 5 个网格共 700 多个小区连

续一个月的数据，以 15 min 为粒度采集测量报告数据，配置参数以天为粒度进行采集，性能数据以 15 min 粒度进行采集，对处理后的格式化数据进行随机划分，按照 8: 2 的比例划分训练集和测试集，现网采集数据预处理结果见表 1。

表 1 现网采集数据的预处理结果

数据类别	数量/条
MR 数据	33 839
配置数据	2 224
性能数据	64 655

3.2 模型训练与结果分析

根据上文提出的小区场景划分方法，使用 MR 数据对小区场景进行划分，场景自动划分采用 K-means 聚类算法，算法的时间复杂度为： $O(t \times k \times m \times n)$ ，空间复杂度为： $O(m \times (n + k))$ ，其中， t 为迭代次数， k 为簇的数目， m 为特征数， n 为样本数。一般 t 、 k 、 m 均可认为是常量，所以时间和空间复杂度可以简化为 $O(n)$ 。对每个类别的 MR 统计数据求均值，各场景类别 MR 统计数据均值分布如图 5 所示。

对所划分的 4 类小区场景分别进行配置参数与性能参数的关系模型训练，利用深度学习算法模型学习配置参数与性能指标参数的内在关系，采用均方误差（MSE）、平均绝对误差（mean absolute error, MAE）以及 R-Squared 对深度学习算法训练得到的模型进行评价。DNN 算法的复杂度主要由隐藏层、隐藏单元以及样本的数量决定。DNN 模型评价指标见表 2，评价结果显示，深度学习算法对数据具有较好的拟合度，预测数据与真实数据的误差较小。

根据训练得到的 DNN 模型进行优化搜索，输出优化配置建议以及调整效果预测结果。将问题转化为爬山算法以解决多维约束的优化问题，输出优化配置建议与调整效果预测结果，见表 3。本文所选取的小区场景为 F1D1D2FDD 900 MHz，

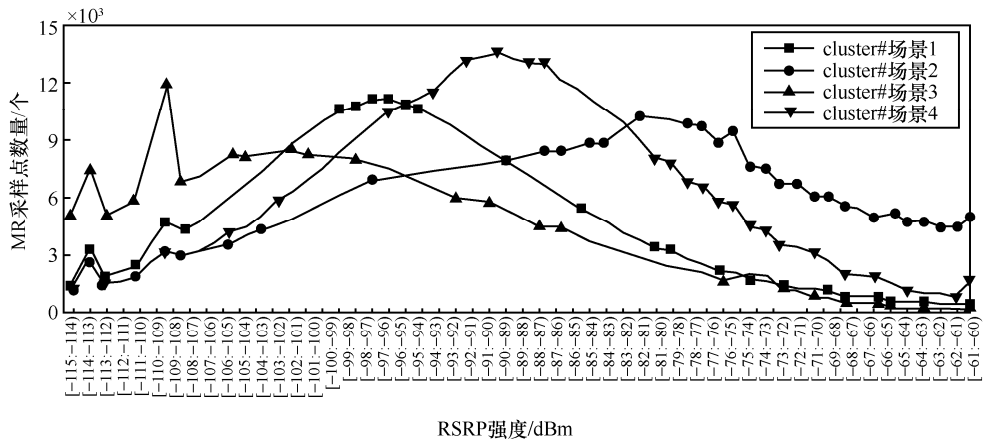


图 5 各场景类别 MR 统计数据均值分布

表 2 DNN 模型评价指标

评价指标	性能参数	A 类场景	B 类场景	C 类场景	D 类场景
MAE	F1	0.050 305 1	0.085 069	0.083 26	0.072
	D1	0.075 814 53	0.081 356	0.090 3	0.097 4
	D2	0.049 532 81	0.062 331	0.059 16	0.063 4
	FDD 900 MHz	0.066 624 23	0.056 063	0.077 29	0.082
MSE	F1	0.004 173 82	0.012 049	0.014 59	0.008 6
	D1	0.009 467 74	0.010 753	0.011 83	0.015 2
	D2	0.004 050 8	0.007 063	0.005 61	0.006 9
	FDD 900 MHz	0.006 955 2	0.004 995	0.009 56	0.011 6
R-squared	F1	0.041 087 96	0.072 958	0.111 81	0.122 8
	D1	0.287 457 44	0.287 098	0.233 28	0.146 1
	D2	0.296 022 1	0.197 456	0.290 42	0.150 5
	FDD 900 MHz	0.162 224 46	0.226 527	0.252 51	0.065 3

表 3 优化配置建议与调整效果预测结果

结果		A 类场景	B 类场景	C 类场景	D 类场景
优化配置参数		-103	-109	-107	-110
		-101	-103	-99	-92
		-87	-93	-86	-101
		-98	-101	-98	-88
		-88	-88	-80	-95
		-108	-85	-84	-97
		...	...	...	...
业务预测比例	F1	0.305	0.311	0.306	0.324
	D1	0.310	0.311	0.312	0.296
	D2	0.307	0.304	0.313	0.311
	FDD 900 MHz	0.077	0.071	0.068	0.067



其中, F1、D1 和 D2 在现网中均为 20 MHz 带宽, FDD 900 MHz 为 5 MHz 带宽, 带宽比 $F1:D1:D2:FDD\ 900\ MHz=0.307\ 7:0.307\ 7:0.307\ 7:0.076\ 9$ 。因此, 4 个频点所承载的业务量越接近带宽比, 则业务均衡度越好。

依托省级大数据平台以及云资源池的资源能力, 以上对于小区场景划分、DNN 模型训练以及最优配置参数求解均在离线端完成, 属于轻量化的运算。在线部署部分的运算效率主要取决于样本数量, 与离线端相比, 复杂度大大减少, 更有益于实时高效的参数优化。目前人工配置方案耗费大量人力和时间, 且负荷不均衡小区的改善比例较低。测试结果显示, 基于机器学习的负荷优化方案可按照小区的带宽容量做到精细化的负荷分担, 负荷不均衡的小区的改善比例达到 90% 以上, 在节省大量人力的同时, 也提高了负荷优化的精确度。

4 结束语

本文提出了一种基于聚类分析和深度学习的多频多模网络负载均衡方案。同时, 联合厂商进行了负荷优化系统的研发以及现网验证, 在选取的测试网格中针对现网负荷优化不佳的小区进行现网测试验证, 测试改善比例可达到 90%。2020 年将扩大测试范围并规模验证方案的可移植性, 进一步优化实施方案以加速负荷优化系统的应用。

本系统真正地将机器学习与无线网络优化结合起来, 形成了基于机器学习的网络参数调优的技术方案, 且有效地解决了 LTE 多频多模网络的负荷不均衡问题。后续将深入量化分析用户分布对负荷优化指标数据的影响, 探索小区粒度与时间粒度更加精细化的动态自适应负荷优化方案。

参考文献:

- [1] 3GPP. Evolved universal terrestrial access network (E-UTRAN); layer 2 measurements: TS36.314 [S]. 2010.
- [2] 郭佳睿, 魏进武, 张云勇. 大数据助力运营商提升规模化运营核心力策略[J]. 电信科学, 2018, 34(1): 120-125.
- [3] 周一青, 李国杰. 未来移动通信系统中的通信与计算融合[J]. 电信科学, 2018, 34(3): 1-7.
- [4] 王志宏, 杨震. 人工智能技术研究及未来智能化信息服务体系的思考[J]. 电信科学, 2017, 33(5): 1-11.
- [5] 许波, 付雷, 张强. LTE 多载波负荷均衡优化浅析[J]. 邮电设计技术, 2018(7): 31-35.
- [6] TIAN S, LI X, JI H. Mobility prediction scheme for optimized load balance in heterogeneous networks[C]//Proceedings of 2018 IEEE Globecom Workshops. Piscataway: IEEE Press, 2018: 1-6.
- [7] XU Y, XU W, WANG Z, et al. Deep reinforcement learning based mobility load balancing under multiple behavior policies[C]//Proceedings of 2019 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2019: 1-6.

[作者简介]



邱亚星(1990—), 女, 中国移动通信有限公司研究院工程师, 主要从事无线大数据、无线网络智能优化技术以及无线网络智能化节能技术的研究与应用工作。

王希栋(1984—), 男, 中国移动通信有限公司研究院工程师, 主要从事无线网络节能技术以及绿色业务、高效网络的协同优化研究工作。

边森(1980—), 男, 中国移动通信有限公司研究院工程师, 主要从事 TD-SCDMA/TD-LTE 绿色通信、新能源基站等领域的研究工作。

岳磊(1981—), 男, 中国移动通信集团广西有限公司高级工程师, 主要从事无线网络规划与优化领域的研究工作。