[数据分析]

基于多源数据分析的业扩报装负荷预测方法

文/陈建民
国网安徽省电力有限公司巢湖市供电公司

提出基于多源数据分析方法，通过灰色关联度方法对业扩报装负荷与影响因素的关联度分析，采用轮廓系数对影响因素的特征进行聚类指标评价分析，对业扩报装的多源数据进行数据采集处理及训练，提出负荷预测方案。实验结果表明，所提出的方法准确度高，误差小，为电客户提供安全、可靠、经济和合理的用电方案建议，进而推动电力领域的稳定发展。

随着经济的快速发展和社会的不断进步，电力能源的需求日益增长，业扩报装业务量也呈现出持续上升的态势。业扩报装作为供电企业与用电客户之间连接的重要环节，关乎着电网的合理规划、稳定运行以及客户用电需求的有效满足。

在业扩报装工作中，负荷预测起着关键作用。准确的负荷预测能够帮助供电企业提前布局电网资源，合理规划电网建设与改造，避免因电力供应不足或过剩而带来的诸多问题。然而，传统的负荷预测方法往往依赖于有限的历史数据或单一数据源，难以全面、精准地把握复杂多变的用电需求情况[1]。

如今，身处大数据时代，多源数据不断涌现，涵盖了诸如历史用电记录、地理信息数据、气象数据、社会经济发展数据以及用户报装申请相关信息等。这些不同来源的数据蕴含着丰富且多元的特征，若能充分挖掘并整合利用，将为业扩报装负荷预测提供更全面、更具深度的视角。

基于此，探索基于多源数据分析的业扩报装负荷预测方法就显得尤为重要。本文旨在深入研究如何有效融合多源数据，并运用灰色关联度算法、模糊C-均值聚类算法、参数训练，提升业扩报装负荷预测的准确性和可靠性，为优化电网规划、提高供电服务质量贡献力量。

业扩报装负荷与影响因素的关联度分析

1. 灰色关联度算法

在电力系统长期负荷预测中，数理统计的相关性分析方法有不足，对数据量、计算量要求高且需特定概率分布对象。而灰色关联度算法优势明显，通过对比变化趋势判断相关程度，不依赖大量数据，也无分布形态要求，适用于电力系统[2]。用来分析电力用户负荷与影响因素关系很恰当，能精准把握关键因素，助力中长期负荷预测、资源配置，保障电力系统稳定高效运行。

2. 灰色关联度分析过程

在业扩报装工作中，准确把握负荷与诸多影响因素之间的关联度对于合理规划电力供应、优化电网布局等有着重要意义。本文采用灰色关联度算法对各类电力用户所属行业的总用电量与多个影响因素之间的相关性进行定量分析，以下是具体的分析过程介绍。

首先选择分析对象。综合考虑各方面因素，本文选取年度地区生产总值、三大产业生产总值、

各类电力用户所属行业生产总值、城乡常住人口总数、城乡人均可支配收入水平及城乡人均消费水平作为影响因素方面的分析对象。将各类电力用户所属行业的历年总用电量作为分析对象的另一方。这些选取的对象涵盖了经济、人口以及消费等多维度，能够较为全面地反映与用电负荷之间可能存在的关联。

然后形成比较序列与参考序列。通过采集分析对象中各影响因素的历年数据，将其整合形成比较序列[3]。设影响因素对应的比较序列共有a个，对于第i个比较序列(i=1,2,⋯,a)，记为x_i，其表示形式为

$$X_i=\{x_i(1),x_i(2),...,x_i(n)\} \tag{1}$$

式中，$x_i(n)$表示第i个影响因素在第n年的数据。

最后进行数据序列归一化处理。由于不同分析对象间的数值因量纲不同，直接相互比较是没有实际意义的，所以需要对各比较序列与各参考序列中的数据进行消去量纲的归一化处理。本文采用首一化的归一化处理方法，其具体公式为

$$\begin{aligned}X_i^1(t)&=\frac{X_i(t)}{X_i(I)}\\X_j^1(t)&=\frac{X_j(t)}{X_j(I)}\end{aligned} \tag{2}$$

式中，$X_j^1(t)$表示第j个行业第t年的数据；$X_i^1(t)$表示第i个影响原始第t年的数据；

数据归一化处理后，为了精细呈现数据变化差异，突出特征，为特征数据分类提供有力支撑，对参考序列进行关联系数矩阵计算，计算公式为

$$\Delta(j,i)=|Y_t^1(t)*\sigma-X_j^1(t)| \tag{3}$$

式中，σ表示分辨系数；$X_j^1(t)$表示第j个参考序列的差序列矩阵中所有元素的最大值和最小值；$X_i^1(t)$表示第j个序列与第i个比较序列的差序列向量。

每个行业的用电数据输出为不同的参考序列向量，计算不同元素影响权重，权重越大，影响因素的关联度越大。关联度计算公式为

$$r_{ji}=\sum_t\omega(t)\lambda_{ij}(t) \tag{4}$$

式中，r_{ji}表示第j个行业用电量与第i个影响因素的关联度；ω表示权重向量；$\lambda_{ji}(t)$表示关联系数矩阵。

通过这样一系列严谨的步骤及相应公式的运用，利用灰色关联度算法精准分析出业扩报装负荷与各影响因素之间的关联程度，从而为电力系统相关决策提供有力的数据支撑[4]。

特征符合分类

1. 多源数据分类过程

本文通过模糊C-均值聚类算法对多源数据进行分类处理，为业扩报装负荷预测提供数据支撑。

生成初始隶属度矩阵[5]。不同用户的负荷存在众多不同的类型，比如商业负荷、工业负荷和军民生活负荷等，隶属度矩阵的目的是计算不同用户所属负荷的程度。矩阵公式为

$$\begin{aligned}&\sum_{i=1}^{c}u_o=1\\&0\leqslant u_o\leqslant 1\\&\sum_{j=1}^{b}u_o\neq 1\end{aligned} \tag{5}$$

式中，u_o表示关联度向量；b表示参数序列向量的隶属度；

欧式距离可衡量数据点相似性，距离近的点更可能属同一类。在聚类、分类模型中常作判断依据，辅助分类操作，还能用于评估分类效果，是数据分类时不可或缺的重要参考指标。本文通过欧式距离完成电力用户的多源数据分类，欧式距离分类公式为

$$V(U,C,R)=\sum_{i=1}^{c}\sum_{j=1}^{b}u_{ij}^2d_{ij}^2 \tag{6}$$

式中，u_{ij}^2表示聚类中心；$V(U,C,R)$表示关联度向量；d_{ij}^2表示关联度向量之间的欧式距离。

2. 聚类结果的评价指标

评估多源数据特征的聚类结果质量时，轮廓系数是一个关键指标。轮廓系数能综合考量簇内紧密程度与簇间分离程度，以此判断聚类效果的优劣。轮廓系数取值范围在[−1，1] 之间。对于样本数据，样本数据同簇之间的平均距离代表数据之间的紧密程度，距离越小，数据紧密度越高。轮廓系数计算公式为

$$s(i)=\frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \tag{7}$$

式中，$a(i)$表示样品到同簇内其他样品的平均距离；$b(i)$表示不同样本之间的平均距离最小值；i表示样本数；

整体聚类的轮廓系数S就是所有样本点轮廓系数的平均值。S值越接近1，表示聚类效果越好，意味

着簇内样本紧密且簇间分离明显，多源数据特征突出；若S接近−1，则聚类效果差。

业扩报装负荷预测方法

1. 数据采集与预处理

在业扩报装负荷预测中，数据采集过往业扩报装的各类详尽数据，涵盖用户基本信息、用电容量、用电性质、报装时间等多方面内容，构建起丰富的数据样本集。

接着是数据处理环节[6]。运用灰色关联度分析，挖掘出与负荷关联性较强的关键因素，比如不同用电性质对负荷变化的影响程度等，筛选出重要变量用于后续预测。

然后采用模糊C均值聚类算法，依据数据特征将众多业扩报装案例聚类分组，使相似情况归为一类，更清晰呈现数据内在规律。

最后借助轮廓系数来评估聚类效果，优化聚类数量等参数，确保分组科学合理。通过这一系列方法协同作用，对采集的数据进行有效处理

2. 参数训练

在业扩报装负荷预测中，精准的参数训练准备是构建可靠预测模型的关键环节，其中涉及到多个关键参数及相应的计算公式来保障训练效果。

针对正则化参数的参数训练方法。为了防止模型过拟合，在损失函数中引入正则化项，常用的岭回归正则化后的损失函数公式为

$$J=\sum_{i=1}^{m}(m_i-\theta^T\times z)+\eta\sum_{j=1}\rho_i^2 \tag{8}$$

式中，m_i表示样本数量；θ表示输入特征向量；z表示模型参数向量；η表示特征维度；ρ表示需要确定的正则化参数。

通过交叉验证等方法不断调整的值，找到使模型在训练集和验证集上性能表现最佳的参数取值，让模型在拟合数据的同时保持良好的泛化能力，为业扩报装负荷预测提供坚实的参数基础[7]。

3. 业扩报装的负荷预测

参数训练后，将原始历年不同行业电力用量数据，分类为训练数据和测试数据两部分，本文采用RBM模型完成业扩报装的负荷预测。RBM模型由多个神经网络组成，层层排列，利用隐含层和可视层双重叠加的方式完成数据的计算和预测[8]。RBM的线性预测模型为

$$P=\sum g_i*x_i+b \tag{9}$$

式中，P表示预测的业扩报装负荷；g_i表示第i个输入的影响因素数据；b表示对应的权重参数，也就是RBM通过训练学习到的该影响因素对于负荷的影响程度权重；x_i表示偏置项，用于整体调整预测结果的基准值。

将训练好的数据输入RBM模型，依据玻尔兹曼分布原理，让模型不断调整权重，通过正向计算隐藏层激活状态以及反向重构可见层数据，反复迭代使得重构误差不断减小，输出对应的负荷预测值。

实验研究

本文以华东地区，某省2010—2020年的业扩报装数据为实例数据，为了避免试验结果单一性，分别对采矿业、制造业、建筑业、电力燃气供应业、交通运输业、商业住宿、房地产、公共事业、农林牧渔业及计算机服务业10个用电行业的用电量进行预测分析，分别对应$Y1$、$Y2\cdots Y10$。本文分别从经济因素、人口因素和生活水平三个维度，选出生产总值、城镇人口数量和城镇销售水平等18个影响因素作为试验数据，依次由$X1$、$X2\cdots X18$表示。

首先通过灰色关联度分析方法对10个行业与18个影响因素进行关联度分析，利用模糊C-均值聚类算法将关联度结果聚类分析，得出10个行业用电的主要影响因素，结果见表1。

表1　各行业主要影响因素分类

类别	主要影响因素
类别1（Y4/Y8/Y10）	X1/X3/X5/X8/X15/X17
类别2（Y1/Y3）	X2/X12/X13/X14
类别3（Y5/Y6）	X1/X4/X7/X9/X11/X16/X18
类别4（Y2）	X2/X8/X12/X13/X14/X15/X17
类别5（Y7/Y9）	X4/X6/X10/X11/X16

将分类结果带入RBM训练模型中，将类别作为预测模型的输出变量，主要影响因素为预测模型的出入变量，依次完成五个类别的数据测试预测如图1所示。

本文通过不同行业扩报装预测的结果和实际值的差数作为评价预测方法精确度的指标。基于多源数据分析的业扩报装负荷预测方法和传统负密度分析方法针对类别1三个行业的数据预测结果见表2。

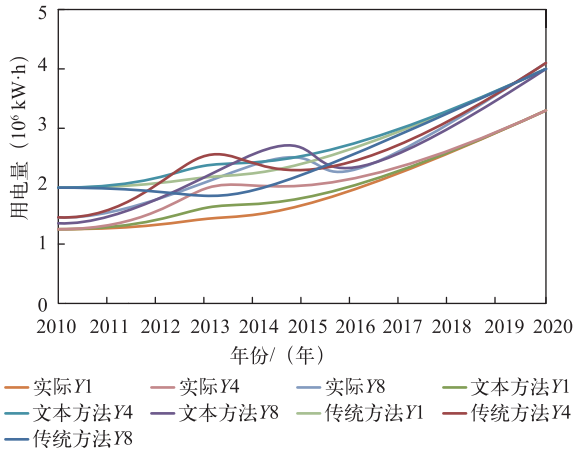

图1　负荷预测结果图

表2　负荷预测误差率情况统计

误差率	项目	本文方法	传统方法
Y4	平均值	2.15	2.35
	最大值	5.24	5.45
	最小值	0.09	0.32
Y8	平均值	0.92	1.44
	最大值	2.38	2.73
	最小值	0.05	0.13
Y1	平均值	2.33	3.52
	最大值	5.63	10.38
	最小值	0.46	0.44
全局	平均值	1.80	2.37
	最大值	5.63	10.38
	最小值	0.09	0.32

将预测结果与实际年限的各行业用电总量进行对比，结果显示，本文提出的基于多源数据分析的业扩报装负荷预测方法对三个行业的负荷预测误差率的平均值、最大值、最小值均小于传统方法，其中本文方法的预测误差最大值为5.63%，误差最小值为0.09%；传统方法误差最大值为10.38%，误差最小值为0.32%，相比之下，本文负荷预测方法具有足够高的精确性。

试验共完成了10个行业的负荷预测，综上本文提出的方法可以适应于各种电力用户的负荷预测，具有更高的实用价值。

结束语

基于多源数据分析的业扩报装负荷预测方法为解决传统预测方式的局限提供了新思路与有效途径。通过灰色关联度算法及多源数据特征分类，整合如历史用电、地理信息、气象以及社会经济等多源数据，借助数据训练算法，更精准地把握业扩报装中的负荷变化趋势。

这不仅有助于供电企业提前做好电网资源调配，科学规划电网建设，增强电网应对用电需求变化的能力，还能进一步提升供电服务质量，更好地满足广大用电客户的需求，推动电力行业与社会经济协同发展。未来，随着数据量的持续增长与分析技术的不断进步，此方法仍有很大的优化空间，期待其在保障电力稳定供应方面发挥更为重要的作用。**EA**

（收稿日期：2025-08-25）

参考文献

[1] 高崇，周鹏，罗强，等．考虑实际负荷增长模式下的配电网供电能力评估方法[J]．南方电网技术，2023，17(7)：115-124．

[2] 黄存强，赵雪，张祥成，等．基于并行数据计算的网格化电网空间负荷预测方法研究[J]．电子设计工程，2023，31(9)：123-127．

[3] 赵紫昱，陈渊睿，陈霆威，等．基于时空图注意力网络的超短期区域负荷预测[J]．电力系统自动化，2024，48(12)：147-155．

[4] 吴靖．基于多源数据分析的区域划分形态识别与优化技术研究[J]．电子设计工程，2023，31(14)：57-61．

[5] 汪超，付晓东，万道春，等．高陡边坡多源数据融合模型构建方法及稳定性分析[J]．矿业研究与开发，2024，44(3)：57-64．

[6] 丁恒春，刘岩，陈雪敏．基于B/S架构的电力客户业扩报装流程智能化管理系统[J]．自动化技术与应用，2023，42(1)：147-151．

[7] 孙云，张云旆，刘立鹏，等．基于滇中引水的TBM数据预处理与特征参数分析[J]．地下空间与工程学报，2023，19(2)：594-608．

[8] 石宏宇，程柯，王信科，等．基于多态数据采集的电力巡检驾驶舱故障预警与辅助决策[J]．机械设计与制造工程，2024，53(6)：95-100．