

基于 K-means 算法的某工业园区经济发展情况监测分析

胡长武 麦晓庆

国网宁夏电力有限公司中卫供电公司 宁夏 中卫 755000

【摘要】：基于 K-means 聚类算法，对某工业园区的经济发展情况进行监测分析，采用大数据挖掘分析技术为工业园区的经济运行监测提供新的参考方向，通过归集工业园区内企业用电量及电费缴纳情况，通过时间、行业分类、用户明细等数据，在园区内经济运行效益的基础上，对各单位用电量进行预测分析，建立工业园区企业经济运行与用电量、电费间的分析评估模型，辅助开展园区内供电服务保障决策，其次通过用户聚类对单个用户或用户群组有进一步的认识，有助于开展工区园区客户经济活动分析和潜在优质客户的整体筛选，为政府发展和经济活动部门开展园区规划、企业信用和债务评价提供参考依据，为电力数据服务工业园区经济发展情况监测分析方面具有创新意义和可推广价值。

【关键词】：工业园区；K-means 算法；经济发展；预测

引言

某工业园区是市委、市人民政府为实现“工业强市”战略目标规划建设的集工业发展、生态绿化为一体的生态园林工业区。

本文通过收集近 5 年某工业园区内 162 家企业用电量及电费缴纳情况，对时间、行业分类、用户明细等维度进行分析企业经济发展情况，并对用电量进行预测^[1]。使用 K-means 算法对用户进行聚类，针对不同用户群体制定相应的应对措施^[2]。通过一系列分析，判断出园区经济发展进展情况，对引入新企业及园区下一步整体规划提供参考。对高价值客户，保障供电质量，保持其稳定；对低价值用户分析其能源使用占比及电量使用过少原因，有针对性的提供解决方案^[3]。

1 问题与挑战

1.1 存在的主要问题

该工业园区存在工业用户数量为 162 户，要完成整个工业园区经济情况的分析，需要对生产规模较大的用户的能源使用结构进行基本分析和收据收集，由于没有类似大数据可提供工业园区客户生产性质的准确材料，需营销业务系统客户性质分析与实地调研相结合，从而较为准确的反映企业能源使用结构，但线下调研收集数据存在数据误差，因此会造成专题分析周期长和效果不佳的问题，同时因电力客户生产用能结构可能较为单一，因此存在局部用能分析不能反映整体经济发展的情况。

1.2 面临的挑战

针对目前面临现状和问题，采取对工业园区覆盖各类企业 162 家客户近 5 年的数据，使用 K-means 算法，根据电费缴纳金额和电费缴纳次数情况对用户聚类后，通过大部分客

户发展情况近似等效分析园区的整体经济发展情况，这本身具有一定的挑战性。

2 分析思路及数据接入

2.1 分析思路

本次分析采用 K-means 算法分析模型，实现工业园区经济发展情况分析主要包括数据收集、数据趋势分析、数据聚类分析、成果分析四个环节。

数据收集：本分析报告所需数据包括某工业园区用户明细、用户近 5 年的电量电费数据。

数据趋势分析：对收集到的数据按时间、行业分类、用户明细、用电量预测等多个维度分析企业用户的用电变化趋势，从而分析其经济发展情况^[4]。

数据聚类分析：使用 K-means 算法对所有用户从电费缴纳金额和次数 2 个维度进行聚类，2 个维度两两组合共聚成 4 类。

成果分析：将企业用户用电量趋势和用户聚类的分析结果汇总，给实际运维部门或政府部门针对园区下一步工作计划提供参考依据。

2.2 数据接入

(1) 数据来源

在供电服务指挥系统中提取某工业园区供电线路所辖的变压器台账；在营销数据库中提取用户明细及近 5 年用户用电量、缴费金额等数据。

(2) 数据字典

本报告中数据共涉及 19 个字段，其中营销系统 13 个、

供服系统6个。数据重要字段如表1所示。

序号	字段名	字段类型	来源	序号	字段名	字段类型	来源
1	配变名称	string	供服系统	11	城衣网	string	营销系统
2	用户编码	string	营销系统	12	地区特征	string	营销系统
3	用户名称	string	营销系统	13	可开放容量	float	供服系统
4	用电类别	string	营销系统	14	额定容量	float	供服系统
5	行业分类	string	营销系统	15	所属台区	string	供服系统
6	所属线路	string	供服系统	16	立户日期	date	营销系统
7	所属变电站	string	供服系统	17	应收年月	date	营销系统
8	维护班组	string	营销系统	18	总电量	float	营销系统
9	运维单位	string	营销系统	19	应收金额	float	营销系统
10	公专变	string	营销系统				

表1 数据处理后的最终字段

3 实现流程及模型构建

3.1 实现流程



图1 实现流程图

(1) 提取变压器明细

在供电服务指挥系统提取变压器台账。

(2) 提取用户明细

使用供服系统中导出的配变名称和线路名称作为筛选范围，在营销数据库中提取用户编号、用户名称、行业分类等数据信息。

(3) 数据处理

提取的用户行业分类是小类，通过行业分类汇总表，统计出其所属大类。

(4) 数据趋势分析

对收集到的数据按时间、行业分类、用户明细、用电量预测等多个维度分析企业用户的用电变化趋势，从而分析其经济发展情况。

3.2 模型构建及主要算法

(1) 算法选取

本文采用 K-means 聚类分析算法，以用户缴费金额和缴费次数2个维度作为依据，将用户分成4类。

(2) 模型搭建

第一步：首先任意选取4个用户，以缴费金额为X轴，缴费次数为Y轴，每个用户都看作是一个点。

第二步：通过欧式距离公式计算剩余点分别到这4个点的距离，选取距离最近的点分成一组。

第三步：计算质心，使用公式，分别求出每组数据的质心。

第四步：计算点到质心的距离，重新计算每个点分别到4个质心的距离，对计算结果重新分组。第二次分组结果为：组1:54个、组2:61个、组3:13个、组4:34个^[5]。

重复第三步和第四步，当最后一次结果与上一次结果一致，说明已经收敛，聚类结算。第n次分组结果为：组1:104个、组2:33个、组3:23个、组4:2个。

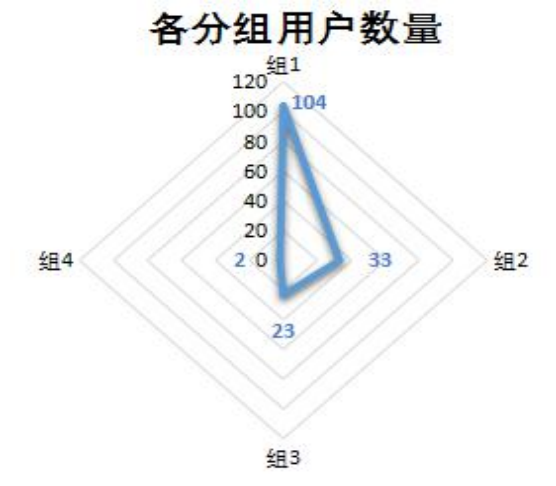


图2 分组结果

对分组结果进行定义，划分成4类用户。

长期稳定优质客户：组3，用户缴费金额高，缴费次数多。

新增优质客户：组4，用户缴费金额高，缴费次数少。

普通客户：组1，用户缴费金额一般，缴费次数多。

不稳定低价值客户：组2，用户缴费金额一般，缴费次数少。

4 监测分析成果

通过 K-means 算法对某工业园区内 162 家用户缴费金额和缴费次数对用户进行聚类, 分别为缴费金额高且缴费次数多的优质客户 23 个、缴费金额高但缴费次数少的新增优质客户 2 个、缴费金额一般但缴费次数多的一般用户 104 个、缴费金额一般且缴费次数少的不稳定用户 33 个。基于数据统计、预测和聚类分析, 对用户电量使用情况和电费缴纳情况进行分析, 反映园区内企业经济发展情况, 有助于对用户电费缴纳及时情况分析, 评估用户欠费风险, 提早预防, 减少用户欠费情况。

5 结语

本文结合工业园区经济发展运行管理实际需求, 分发挥 K-means 算法计算优势与数据挖掘优势, 深入开展了基于大数据的工业园区经济发展监测分析, 运用大数据分析技术, 通过归集时间、行业分类、用户明细等数据, 对用户电量使用情况和电费缴纳情况进行分析, 将园区用户进行分类^[6]。通过用户聚类对单个用户或用户群组有进一步的认识, 有助于开展工业园区客户经济活动分析和潜在优质客户的整体筛选, 为政府发展和经济活动部门开展园区规划、企业信用和债务评价提供参考依据。

参考文献:

- [1] 赵腾,王林童,张焰,田世明.采用互信息与随机森林算法的用户用电关联因素辨识及用电量预测方法[J].中国电机工程学报,2016,36(03):604-614.
- [2] 魏小曼,余昆,陈星莺,颜拥,张爽,景伟强.基于 Affinity propagation 和 K-means 算法的电力大用户细分方法分析[J].电力需求侧管理,2018,20(01):15-19+35.
- [3] 赵晨,尹玉芬.电力体制改革下工业用电客户差异化管理策略[J].电力与能源,2018,3(05):708-713.
- [4] 王利利,张琳娟,许长清,武宏波,李翼铭,黄玉晶.能源互联网背景下园区用户画像及成熟度评价模型研究[J].中国电力,2020,53(08):19-28.
- [5] 尚海昆.K-means 聚类算法的研究[D].华北电力大学(河北),2009.
- [6] 凌德祥,黄拓,关晓林,赵严.基于大数据的电力客户行为分析体系研究及实践[J].电力大数据,2018,21(10):13-17.