

# 基于大数据分析的县公司综合评价策略

薛振宇<sup>1</sup>, 胡航海<sup>2</sup>, 宋毅<sup>1</sup>, 吴志力<sup>1</sup>, 刘道新<sup>2</sup>, 冯昊<sup>3</sup>, 宋红芳<sup>3</sup>

(1. 国网北京经济技术研究院, 北京 102209; 2. 国家电网公司, 北京 100052;

3. 国网浙江省电力公司, 浙江 杭州 310007)

**摘要:** 提出一种基于大数据分析的县公司综合评价策略。以 1780 家县公司数据为基础, 建立考虑统计分析、决策树分析、数据包络分析等大数据技术的指标挖掘策略; 通过信息挖掘构建一套适用于县公司综合评价的指标体系。以经济因素和企业因素为规则, 建立基于县公司聚类的差异化权重计算方法, 实现对县公司差异化发展内生和外生因素的联合判别, 克服综合评价结果解释的单一化。通过浙江省的实证化应用, 验证了所提策略的有效性和实用性。

**关键词:** 大数据分析; 县公司; 综合评价; 指标体系; 数据挖掘

**中图分类号:** TM 761

**文献标识码:** A

**DOI:** 10.16081/j.issn.1006-6047.2017.09.028

## 0 引言

县级供电企业是配电网建设与运行的主体, 企业的经营能力对配电网的发展具有决定性作用。与城市地区的电网评价工作<sup>[1-2]</sup>相比, 县公司所处地区的经济社会、人文环境、地理气象等条件差异很大, 给综合评价的可比性和合理性带来诸多挑战, 相关的文献资料鲜有报道。随着电网的精细发展, 如何通过综合评价合理定位县公司发展方向, 对引导电网发展和企业经营具有十分重要的现实意义。

大数据分析 BDA(Big Data Analysis)<sup>[3-4]</sup>技术作为因果推导分析的重要补充工具, 近年来得到广泛应用<sup>[5-6]</sup>。在需求侧管理<sup>[7-8]</sup>方面, BDA 深度挖掘电力用户侧数据, 辅助实现用电管理及更准确的短期负荷预测; 在状态检测<sup>[9-10]</sup>方面, BDA 帮助完成状态数据检测和快速分析处理, 实现对输配电设备状态评估与检测; 在电力系统仿真<sup>[11-12]</sup>方面, BDA 可辅助给出系统稳定研判和防控告警。对县公司而言, 其具有企业数量多、电网数据多、数据结构多的大数据特征。同时, 受经济社会、管理理念、历史基础等多方面因素影响, 企业发展差异巨大。相关统计显示, 全国县公司售电量最大的企业是最小的 3000 多倍, 供电可靠率最高的企业是最低的上万倍。因此, 县公司综合评价需要处理好海量数据和发展差异 2 个关键难题。

为此, 本文提出一种基于数据挖掘分析的县公司综合评价方法。以 1780 家县公司企业数据为基础构建评价指标体系, 发展出基于统计分析、决策树 DT(Decision-Tree)分析、数据包络分析 DEA(Data Envelopment Analysis)等大数据技术的核心指标提炼策略, 规避了传统方法中指标选取的主观性。充分考虑企业发展差距, 提出基于系统聚类技术的差

异化分类思路, 形成县公司类内相比、类间参考的管理策略。通过对浙江省 64 家县公司发展内生因和外驱动的实用化应用, 证明了所提方法的有效性和实用性。

## 1 基于 BDA 的综合评价策略

### 1.1 基于 BDA 的评价指标集构建

县公司企业数据资源庞大(1780 家县公司样本), 具有指标类型多、内在关系复杂、地区差异悬殊的特点。其内部数据既包括电网规划、建设、运行、营销等业务部门数据, 也包括财务、人资等管理部门数据; 外部数据以地方经济、社会、人口、气象、环境等数据为主。通过 BDA, 可从海量、低价值密度的数据中挖掘核心指标, 形成企业决策指标集合, 主要流程见图 1。

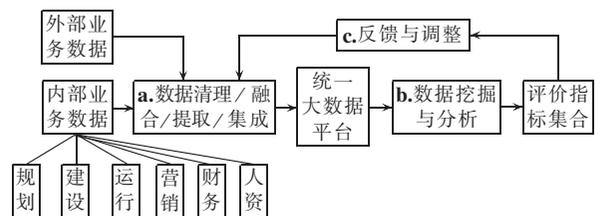


图 1 县公司评价指标 BDA 过程

Fig.1 County company evaluation index set generation based on big data analysis

**a.** 通过数据清理、数据融合、数据提取、数据集成等技术, 形成统一的大数据平台。数据清理可消除不一致数据, 如校验营销系统与运行系统的负荷电量数据; 数据融合可实现同一设备在不同系统中属性的整合, 如整合台账信息与运行状态; 数据提取对研究所需指标针对性抽取相关数据; 数据集成将多种数据集中组合, 满足后续分析需要。

**b.** 通过 BDA, 建立关键评价指标数据集合。选取数理统计、聚类分析、DT 分析、DEA 效率分析等

BDA 方法中的某一种,对若干指标的全量数据(1780 个)进行挖掘,再对挖掘得到的知识进行评估和判断,从而构造出评价指标集合。

c. BDA 高度依赖数据,在步骤 a 的数据清理过程中往往不能完全剔除无效或错误数据,步骤 b 的 BDA 能够进一步对数据进行甄别,提高数据质量。对偏离常规认知的评价指标,须重新处理并进行 BDA,经过若干次循环调整,最终得到一套完备的核心评价指标集合。

## 1.2 基于 BDA 的综合评价方法

在综合评价的指标选取中引入 BDA 技术,既可辅助形成基于知识判断的评价指标,改善传统综合评价指标选取的主观性,也可对县公司进行差异化评价,更准确地剖析企业薄弱环节。基于 BDA 的县公司综合评价过程如图 2 所示。

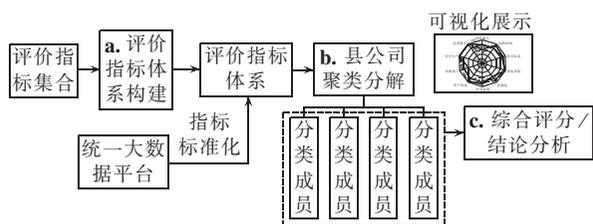


图 2 基于 BDA 的县公司综合评价

Fig.2 Comprehensive county company evaluation based on big data analysis

a. 在知识挖掘的评价指标集合基础上,综合专家经验判断,构建一套统一的综合评价指标体系,并采用基于德尔菲法的改进层次分析方法,为各指标设定综合评分的权重系数。

b. 考虑企业发展差异,引入聚类分析方法,按照一定的分类规则将县公司分为不同类别,不同类型之间主要比较整体差距,同类型内的县公司主要比较具体差异。

c. 结合可视化技术,对各类型县公司的综合评分结果进行可视化展示,挖掘同一类型县公司的共同特征,比较不同类型之间的发展差距,为县公司发展提供决策建议。

## 2 基于 BDA 的县公司评价核心指标挖掘

### 2.1 基于统计分析的指标挖掘技术

(1)相关性分析 CA(Correlation Analysis)。依托 CA 对指标进行层层深入,从关联密切的若干个指标中提取得到具有代表性的核心指标。以 GDP 与电量的 CA 研判为例:经济社会发展对电量影响显著,GDP 与全社会用电量的相关系数为 0.809,与售电量的相关系数为 0.821。由于全社会用电量中计及了厂用电、线损电量等因素,这些电量与 GDP 的关系较弱,因此 GDP 对售电量的作用更为突出。从图 3 所示

的 GDP 与售电量的相关关系散点图来看,1780 个县公司的 GDP 与售电量基本呈直线趋势分布,特别是在浙江、江苏等工业发达省份,两者的相关系数达到 0.9 以上,呈强相关。因此,评价选取售电量作为反映企业受经济社会发展推动效果的核心指标。

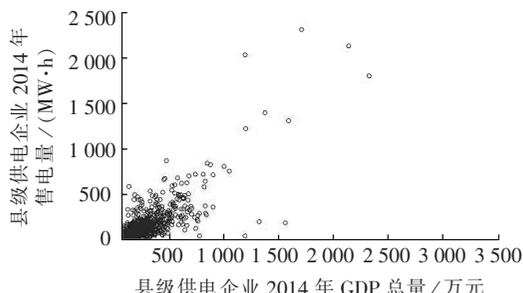


图 3 售电量与 GDP 相关性关系散点图

Fig.3 Scatter diagram of electricity sale vs. GDP

CA 适合分析县公司较为宏观的指标,相关结果显示:电网公司作为公益性企业,其售电量与利润总额相关系数为 0.552,彼此替代性不强;综合线损率与利润总额的相关系数为 -0.128,即控制线损有利于提升利润总额,但 2 个指标彼此替代性不强。因此,售电量、利润总额、综合线损率均作为核心评估指标。

(2)主成分分析 PCA(Principal Component Analysis)。依托 PCA 对多个指标提取,按主成分贡献度选取 1~2 个指标作为代表。以县公司财务管理数据的 PCA 研判为例:18 个主要财务指标的解释方差在主成分为 5 时,累计效果达 84.501%,即 18 个指标可浓缩为 5 个左右。成分矩阵的载荷系数见表 1。

指标的载荷越大表明该指标与对应主成分的关系

表 1 资产财务指标的成分矩阵

Table 1 Composition matrix of capital and financial indexes

指标	成分矩阵				
	成分 1	成分 2	成分 3	成分 4	成分 5
资产总额	0.930	0.157	0.178	-0.029	-0.085
流动资产	0.538	0.114	0.196	-0.313	-0.598
非流动资产	0.885	0.140	0.133	0.089	0.134
负债总额	0.636	0.579	0.459	-0.022	0.040
流动负债	0.678	0.165	0.552	-0.041	-0.205
非流动负债	0.182	0.801	0.041	0.018	0.359
资产负债率	-0.289	0.534	0.354	0.176	0.243
企业资产占比	0.200	-0.367	0.043	-0.196	0.505
主营业务收入	0.907	-0.192	-0.059	-0.045	0.017
主营业务成本	0.928	-0.226	-0.048	-0.039	0.025
利润总额	0.486	0.470	-0.690	-0.153	-0.049
净利润	0.425	0.509	-0.704	-0.148	-0.029
可控费用	0.801	-0.184	0.093	0.044	0.188
中间投入	0.888	-0.230	-0.033	-0.190	0.046
应交增值税	0.587	-0.165	-0.106	0.689	-0.056
工业总产值	0.938	-0.196	-0.081	-0.045	0.020
工业增加值	0.636	0.001	-0.207	0.617	-0.092
农网维护管理费	0.630	-0.252	-0.019	-0.256	0.283

关联性越强。第 1 个主成分反映企业资产总量,有 11 个备选指标;第 2 个反映企业负债,有 2 个备选指标;第 3 个反映利润,有 2 个备选指标;第 4 个反映纳税,有 1 个备选指标;第 5 个反映资产构成,有 2 个备选指标。结合县公司经营特点,选定主营业务成本、资产负债率、利润总额、应交增值税、企业资产占比 5 个指标。

PCA 适用于对同一专业管理范畴内的数据进行分析,如财务数据、经济社会数据、电网规模数据等。相关结果显示:对变压器、线路负载运行指标进行 PCA 筛选得到容载比、户均配变容量、主变 N-1 通过率、线路 N-1 通过率和 10 kV 台区重载比例 5 个指标,作为电网发展的核心指标。

### 2.2 基于 DT 的指标挖掘技术

依托 DT<sup>[13]</sup>进一步挖掘反映电网主要特征的核心指标。以 10 kV 电网指标的 DT 研判为例:以 PCA 确定的户均配变容量作为 DT 因变量,将 10 kV 线路平均供电半径、10 kV 线路重载比例、10 kV 线路轻载比例、10 kV 台区重载比例、10 kV 台区轻载比例、10 kV 线路 N-1 通过率、10 kV 线路重损比例、10 kV 台区重损比例、低电压台区共 9 个反映中压配电网运行的指标为自变量,按 CHAID 增长进行 DT 分析。得到最大树深度为 3 的分类树,如图 4 所示(部分企业数据缺失,样本数量 1659)。则由户均配变容量、线路平均供电半径和 10 kV 台区重损比例 3 个指标即可构成中压配电网的主要评估指标。

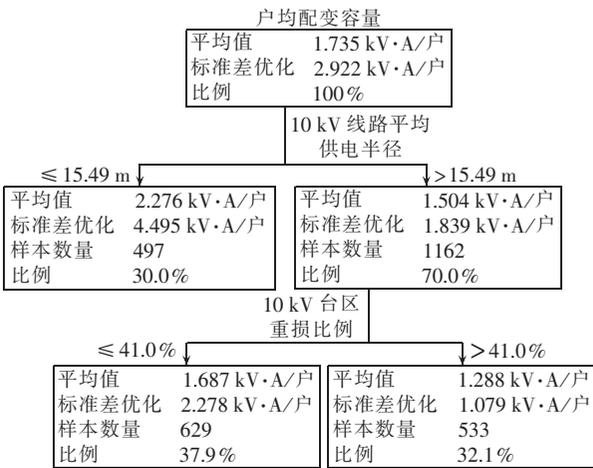


图 4 10 kV 电网运行特征指标 DT 分析  
Fig.4 Decision-tree analysis of 10 kV-grid operational indexes

DT 适合以统计分析所得指标为源,进一步追溯体现县公司差异规律的关键指标。在上述 10 kV 电网运行指标的基础上,10 kV 台区重损比例可再向下分解出综合电压合格率和用电信息采集系统覆盖率。

### 2.3 基于 DEA 方法的指标挖掘技术

依托 DEA<sup>[14]</sup>给出提升企业相对效率的指标改善空间,提炼影响企业未来发展的核心指标。以县公

司投入产出的 DEA 研判为例:在随机选取的 84 家测试样本中,DEA 效率小于 1 的有 11 家,投入和产出对应的优化空间分别如表 2、表 3 所示。

表 2 DEA 评估投入指标改进方案  
Table 2 Input index improvement by DEA evaluation

样本	效率值	购售价差/ [元·(kW·h) <sup>-1</sup> ]	营销投入/万元	研究/开发费/万元	生产/大修/万元	教育/培训/万元	户均配变容量/ (kV·A·户 <sup>-1</sup> )
1	1	0	0	-92	-341	-6	-2.37
2	1	0	-257	-9	-545	0	0
3	1	0	0	0	-279	-1	0
4	1	0	-39	-29	-331	0	-2.78
5	1	0	0	-4	-272	-19	0
6	1	0	-640	-53	0	0	-0.82
7	1	0	0	-128	0	0	0
8	1	0	0	-20	0	-20	0
9	1	0	0	0	-458	0	-0.57
10	1	-103	-390	0	0	0	0
11	1	0	0	0	-334	0	-1.58

表 3 DEA 评估产出指标改进方案  
Table 3 Output index improvement by DEA evaluation

样本	效率值	线损率/%	供电可靠率/%	全员劳动生产率/%	利润总额/万元	资产负债率/%
1	0.78	0	0.28	26	0	0
2	0.89	0	0	17	315	4
3	0.99	1.44	0	0	1420	13
4	0.88	0.23	0	58	3448	0
5	0.72	0	0	47	3129	0
6	0.96	0	0	14	867	0
7	0.74	0	0.21	5	0	0
8	0.94	1.70	0	13	405	12
9	0.85	0	0	45	373	0
10	0.87	1.99	0	22	331	9
11	0.90	0.95	0	12	588	0

投入上,降低户均配变容量和控制生产大修成本能有效提高效率,降低户均配变容量可认为是控制基建投入;产出上,提升全员劳动生产率和利润总额能有效提高效率。比较表 2、3 可知,投入优化可将效率值提升至 1,产出优化均未达到 1。故评价时权重系数设定应当倾向投入指标。

## 3 基于 BDA 的县公司综合评价方法

### 3.1 综合评价指标体系构建

首先,构建一套县公司电网发展和生产经营的综合评估指标体系。综合 BDA 知识研判和专家意见,电网发展分为供电能力、电网结构、技术运行、供电质量、智能化 5 个二级指标和 14 个三级指标;生产经营分为经营业绩、资产质量、发展投入、安全生产、供电服务 5 个二级指标和 14 个三级指标。评价指标及权重系数分别见表 4、表 5。表中权重系数由基于德尔菲法的改进层次分析<sup>[15]</sup>方法确定。

为了清晰表达评价维度,将反映企业某一特征的

表 4 县公司电网发展评价指标体系

Table 4 Evaluation index system of grid development for county company

一级指标	二级指标	三级指标	权重系数	
电网发展	供电能力 0.251	容载比	0.621	
		户均配变容量	0.379	
	电网结构 0.276	10 kV 线路平均供电半径	线路 N-1 通过率	0.348
			主变 N-1 通过率	0.424
			主(配)变运行年限	0.289
			线路运行年限	0.275
	装备运行 0.118	10 kV 台区重载比例	10 kV 台区重载比例	0.271
			10 kV 台区重损比例	0.165
			供电质量	0.243
	智能化 0.112	用电信息采集系统覆盖率	供电可靠率	0.741
			综合电压合格率	0.286
			智能电表覆盖率	0.315

注:第 2 列数据表示相应二级指标权重系数,后同。

表 5 县公司生产经营评价指标体系

Table 5 Evaluation index system of production and management for county company

一级指标	二级指标	三级指标	权重系数	
生产经营	经营业绩 0.182	综合线损率	0.355	
		主营业务成本	0.220	
		利润总额	0.216	
		全员劳动生产率	0.208	
	资产质量 0.119	单位资产所供电量	资产负债率	0.386
			单位资产所供负荷	0.295
			单位资产所供负荷	0.319
	发展投入 0.134	单位投资增供电量	单位投资增供电量	0.546
			单位投资增供负荷	0.454
	安全生产 0.336	人身事故伤亡	人身事故伤亡	0.636
			电网事故	0.275
			设备事故	0.089
	供电服务 0.228	95598 工单数	95598 工单数	0.305
			95598 服务满意率	0.695

若干三级指标归纳为一类,形成二级指标。其中,供电能力用于表征电网对负荷的运行承载水平;电网结构用于表征电网的坚强程度;装备运行用于表征主要电气设备性能;供电质量用于表征电网的供电效果;智能化用于表征智能技术的应用程度;经营业绩用于反映企业的经营状况;资产质量用于反映企业资产的运营效率;发展投入用于反映企业投入产出效果;安全生产用于反映企业安全管理水平;供电服务用于反映企业对用户的响应质量。

### 3.2 基于系统聚类的县公司分类方法

其次,将县公司按照一定的规则分成若干可比的类簇。分类规则主要由经济社会条件和电网规模确定,即考虑外部发展条件和内部建设基础。对经济社会条件进行 PCA,提取 GDP 总量、人均 GDP、第三产业占比、城镇化率、供电面积 5 个指标。对电网规模进行 PCA,提取出主变容量和低压线路长度 2 个指标。上述 7 个指标作为县公司分类依据。

具体步骤为:依托系统聚类方法<sup>[16]</sup>中的离差平方方法,即先将  $n$  家县公司各看成一类,每次按照使离差平方和增加最小的 2 类合并,直至所有样本归为一类。系统聚类方法能给出  $n$  个样本逐步聚集为 1 类的过程,按照推荐结果将县公司分为  $k$  类。

### 3.3 评价指标权重差异化处理

最后,根据每一类内多家县公司的整体特征,对表 4、表 5 的权重系数进行修正,形成差异化评价策略。计算中,三级指标权重统一由表 4、表 5 确定,二级指标权重按其得分与预期目标的差距打分后,再对其进行修正确定综合得分。具体步骤如下。

a. 计算  $n$  家县公司的三级指标得分,根据三级指标权重,得到  $n$  家县公司的二级指标得分,公式为:

$$S_i^t = \sum_{j=1}^J \omega_i^{t-j} \times p_i^{t-j} \quad (1)$$

其中, $S_i^t$  为第  $i$  家县公司的第  $t$  个二级指标; $\omega_i^{t-j}$  和  $p_i^{t-j}$  分别为第  $i$  家县公司的第  $t$  个二级指标的第  $j$  个三级指标的权重和得分; $J$  为第  $i$  家县公司的第  $t$  个二级指标的三级指标总数。

b. 对各二级指标按 1— $k$  的评分维度对  $k$  个类打分,构造类间评分矩阵。按行归一化计算,公式为:

$$\alpha_m^t = \frac{x_m^t}{\sum x_m^t} \quad (2)$$

其中, $\alpha_m^t$  为第  $m$  个类第  $t$  个二级指标修正系数; $x_m^t$  为第  $m$  个类第  $t$  个二级指标对应的打分值。

c. 对  $k$  个类的二级指标权重进行修正:

$$\tilde{\omega}_m^t = \omega_m^t \times (\alpha_m^t + \beta) \quad (3)$$

其中, $\tilde{\omega}_m^t$  为第  $m$  个类第  $t$  个二级指标修正后的权重系数; $\omega_m^t$  为第  $m$  个类第  $t$  个二级指标的原始权重系数; $\beta$  为小于 1 的常量,对  $\alpha_m^t$  进行调节,本文取  $\beta=0.5$ 。

d. 对  $\tilde{\omega}_m^t$  进行归一化处理,分别计算每家县公司的综合得分。

## 4 案例应用

对浙江省 64 家县公司进行综合评价。指标体系按照表 4 和表 5 选取,经系统聚类分析确定将 64 家单位分为 5 类。分类结果和各类县公司综合得分情况如表 6 所示,雷达图如图 5 所示。

第 I 类萧山区,是省会杭州市市辖区,经济发展快速,电网规模和资产体量远大于其他单位,但呈现一定的饱和性,综合得分反而不高。未来,其电网建设的重点是强化电网结构,提高供电可靠性,但建设投资带来的回报并不显著。

第 II 类包括 10 家单位,综合评价得分最高。企业的外部经济、社会、人口等方面优势明显,产业园区较为密集,经济发展迅速,处于快速发展期。未来,提高电网供电能力和改善设备重载问题的电网投

表6 县公司综合评价结果  
Table 6 Results of comprehensive county company evaluation

类别	地区	平均分
I	萧山区	87.73
II	慈溪市、桐乡市、瑞安市、富阳市、鄞州区、海宁县、余杭区、义乌市、诸暨市、余姚市	90.17
III	长兴县、温岭市、乐清市、上虞市、苍南县、永康市、东阳市、玉环县、嘉善县、兰溪市	89.79
IV	台州路桥区、平湖市、奉化县、永嘉县、建德市、台州椒江区、海盐县、宁海县、江山市、缙云县、桐庐县、嵊州市、浦江县、武义县、安吉县、象山县、龙游县、新昌县、临海市、黄岩区、德清县、临安市、平阳县	87.40
V	泰顺县、文成县、洞头县、松阳县、嵊泗县、磐安县、开化县、龙泉市、岱山县、景宁县、庆元县、云和县、遂昌县、仙居县、淳安县、常山县、青田县、丽水莲都区、天台县、三门县	84.70

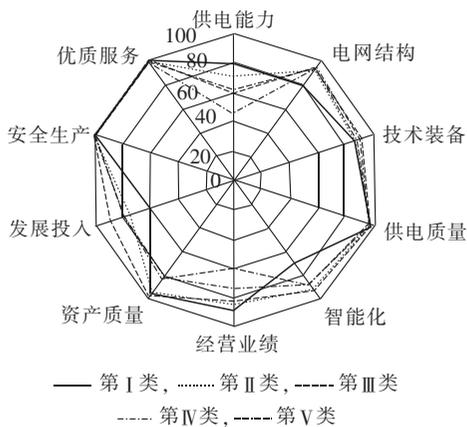


图5 不同县公司分类评价结果雷达图

Fig.5 Radar chart of evaluation results for different county company clusters

资,能够获得较好的经营回报。

第Ⅲ类包括10家单位,综合评价排名第二。地区经济对企业发展有推动作用,但发展不均衡,电网运行轻重载现象并存。未来,需跟踪好经济发展重点,提高电网的有效投资。

第Ⅳ类包括23家单位,与第Ⅰ类得分接近。由于经济社会发展速度较慢,对企业产生一定的限制作用,不同程度地存在电网建设超前,如电网结构、技术装备、供电质量等方面优良,但用电需求有限导致经营绩效偏弱。未来,需控制好电网投资建设进度,预防过度超前。

第Ⅴ类包括20家,排名最后。由于经济社会发展缓慢,且多处山区、丘陵地带,人口密度较低,整体制约了企业发展。既存在容载比偏低引起的设备重载,也存在容载比偏高导致的轻重载并存。生产经营的整体水平较低,综合线损率、利润总额、全员劳动生产率等指标均不理想。未来,需确立合理的电网建设标准,重点满足地区基本生产生活需要。

应用差异化评价,有效解决了因边界条件不同导致企业无法合理对比的问题。浙江的应用表明,所提方法既挖掘出不同类的整体矛盾,发现经济社会外部因素对企业发展具有重要引擎作用;同时实现同类内不同县公司比较,找出与自身条件相近具备可比的参照目标,提升改进。

## 5 结论

以1780家县公司为样本,提出基于统计分析、DT分析、DEA等大数据技术的数据挖掘新思路,构建出能够面向1780家县公司应用的综合评价指标体系。充分考虑到县公司的发展差异和所在地区的经济差异,建立基于系统聚类分析技术的差异化分类方法,将具备可比性的企业分为一类,便于彼此对照比较,不同类型之间进行外在驱动因素分析,使得综合评价的结论更为全面客观。所提方法在浙江省64家县公司实践应用,对发现电网发展薄弱环节,把握企业经营重点,发挥了重要的现实指导作用。

## 参考文献:

- [1] 肖峻,崔艳妍,王建民,等. 配电网规划的综合评价指标体系与方法[J]. 电力系统自动化,2008,32(15):36-40.  
XIAO Jun,CUI Yanyan,WANG Jianmin,et al. A hierarchical performance assessment method on the distribution network planning [J]. Automation of Electric Power Systems,2008,32(15):36-40.
- [2] 韩震焘,黄志伟,葛少云,等. 城市配电网综合评价体系[J]. 电网技术,2012,36(8):95-99.  
HAN Zhentao,HUANG Zhiwei,GE Shaoyun,et al. A comprehensive evaluation system of urban distribution network[J]. Power System Technology,2012,36(8):95-99.
- [3] GANTZ J,REINSEL D. Extracting value from chaos[C]//Proceedings of IDC iView. Framingham,USA:[s.n.],2011:1-12.
- [4] 徐青山,王文帝,林章岁,等. 面向行业大数据特征挖掘的电力经理指数指标体系的建立与应用[J]. 电力自动化设备,2015,35(7):15-21.  
XU Qingshan,WANG Wendi,LIN Zhongsui,et al. Establishment and application of EMI indicator system orienting to massive industrial data mining [J]. Electric Power Automation Equipment,2015,35(7):15-21.
- [5] 宋亚奇,周国亮,朱永利. 智能电网大数据处理技术现状与挑战[J]. 电网技术,2013,37(4):927-935.  
SONG Yaqi,ZHOU Guoliang,ZHU Yongli. Present status and challenges of big data processing in smart grid[J]. Power System Technology,2013,37(4):927-935.
- [6] 张东霞,苗新,刘丽平,等. 智能电网大数据技术发展研究[J]. 中国电机工程学报,2015,35(1):2-12.  
ZHANG Dongxia,MIAO Xin,LIU Liping,et al. Research on development strategy for smart grid big data[J]. Proceedings of the CSEE,2015,35(1):2-12.
- [7] WANG Peijian. Dynamic data center operations with demand-responsive electricity prices in smart grid[J]. IEEE Transactions on Smart Grid,2012,3(4):1743-1754.
- [8] 王德文,孙志伟. 电力用户侧大数据分析与并行负荷预测[J]. 中

- 国电机工程学报,2015,35(3):527-537.
- WANG Dewen,SUN Zhiwei. Big data analysis and parallel load forecasting of electric power user side[J]. Proceedings of the CSEE,2015,35(3):527-537.
- [9] 严英杰,盛戈峰,陈玉峰,等. 基于大数据分析的输变电设备状态数据异常检测方法[J]. 中国电机工程学报,2015,35(1):52-59.
- YAN Yingjie,SHENG Gehao,CHEN Yufeng,et al. A method for anomaly detection of state information of power equipment based on big data analysis[J]. Proceedings of the CSEE,2015,35(1):52-59.
- [10] 张友强,寇凌峰,盛万兴,等. 配电变压器运行状态评估的大数据分析方法[J]. 电网技术,2016,40(3):768-773.
- ZHANG Youqiang,KOU Lingfeng,SHENG Wanxing,et al. Big data analytical method for operating state assessment of distribution transformer[J]. Power System Technology,2016,40(3):768-773.
- [11] 刘道伟,张东霞,孙华东,等. 时空大数据环境下的大电网稳定态势量化评估与自适应防控体系构建[J]. 中国电机工程学报,2015,35(2):268-276.
- LIU Daowei,ZHANG Dongxia,SUN Huadong,et al. Construction of stability situation quantitative assessment and adaptive control system for large-scale power grid in the spatio-temporal big data environment[J]. Proceedings of the CSEE,2015,35(2):268-276.
- [12] 黄天恩,孙宏斌,郭庆来,等. 基于电网运行仿真大数据的知识管理和超前安全预警[J]. 电网技术,2015,39(11):3080-3087.
- HUANG Tianen,SUN Hongbin,GUO Qinglai,et al. Knowledge management and security early warning based on big simulation data in power grid operation[J]. Power System Technology,2015,39(11):3080-3087.
- [13] 白建社,樊波,黄文华,等. 基于决策树的变电站故障诊断知识表示与获取[J]. 电力系统及其自动化学报,2004,16(2):5-8.
- BAI Jianshe,FAN Bo,HUANG Wenhua,et al. Knowledge representation and acquisition based on decision tree for substation fault diagnosis[J]. Proceedings of the CSU-EPSC,2004,16(2):5-8.
- [14] CHIEN C F,LO F Y,LIN J T. Using DEA to measure the relative efficiency of the service center and improve operation efficiency through reorganization[J]. IEEE Transactions on Power Systems,2003,18(1):366-373.
- [15] 王守相,张一帆,葛磊蛟. 新型城镇配电网影响因素解释结构模型[J]. 电力自动化设备,2015,35(11):75-81.
- WANG Shouxiang,ZHANG Yifan,GE Leijiao. Interpretative structural model of influencing factors for distribution network of newtype town[J]. Electric Power Automation Equipment,2015,35(11):75-81.
- [16] 李亦言,严正,冯冬涵. 考虑城市化因素的中长期负荷预测模型[J]. 电力自动化设备,2016,36(4):54-61.
- LI Yiyang,YAN Zheng,FENG Donghan. Mid/long-term load forecasting model considering urbanization characteristics[J]. Electric Power Automation Equipment,2016,36(4):54-61.

#### 作者简介:



薛振宇

薛振宇(1983—),男,山西大同人,高级工程师,博士,主要研究方向为电力系统规划设计、新能源发电、电力系统稳定与控制等(E-mail:xuezhenyu@chinasperi.sgcc.com.cn);

胡航海(1985—),男,福建泉州人,高级工程师,硕士,主要研究方向为电网规划、电力数据分析、企业经营管理等;

宋毅(1977—),男,天津人,高级工程师,博士,主要研究方向为配电网规划、新能源与电动汽车接入、配电系统仿真、电力系统信息化与大数据应用等。

## Comprehensive evaluation based on big data analysis for county electric power company

XUE Zhenyu<sup>1</sup>,HU Hanghai<sup>2</sup>,SONG Yi<sup>1</sup>,WU Zhili<sup>1</sup>,LIU Daoxin<sup>2</sup>,FENG Hao<sup>3</sup>,SONG Hongfang<sup>3</sup>

(1. State Grid Power Economic Research Institute,Beijing 102209,China;2. State Grid Corporation of China, Beijing 100052,China;3. State Grid Zhejiang Electric Power Company, Hangzhou 310007,China)

**Abstract:** A strategy of comprehensive evaluation based on big data analysis is proposed for county electric power company. With the big data technologies,including statistical analysis,decision-tree analysis,data envelopment analysis,etc.,a strategy of index mining is developed based on the data of 1780 county electric power companies and a comprehensive evaluation index system for county electric power company is established by the information mining. With the economic and business factors as the rules,the country electric power companies are clustered and a method of differentiated weight calculation is derived to jointly identify the endogenous and exogenous factors of differentiated county company developments for avoiding the simplified explanation of comprehensive evaluation results. The application of the proposed strategy in Zhejiang Province verifies its effectiveness and feasibility.

**Key words:** big data analysis; county electric power company; comprehensive evaluation; index system; data mining