

基于大数据的电力用户群体识别与分析方法研究

许长清¹, 赵华东², 宋晓辉³

(1. 国网河南省电力公司 经济技术研究院 河南 郑州 450052; 2. 郑州大学 机械工程学院 河南 郑州 450001; 3. 河南省科学院 应用物理研究所 河南 郑州 450008)

摘要: 随着能源互联网及智能用电技术的发展,深入了解电力用户群体特征,并提供精准电力服务,成为“互联网+”智慧能源的重要研究内容.通过对售电体积累的用户社会属性、用电行为等大数据进行分析,构建用户间相似度权重模型,提出基于“用户-标签”二元网络的相似群体识别方法,并通过群体分析获取群属性和典型负荷特征,预测新入网用户的用电行为.该方法通过对标签数据的分析,便于发现群体中的重要用户,为互联网售电体实施个性化用电服务和增值服务推荐提供支撑,进而提升电力服务质量.

关键词: 大数据; 电力用户; 相似群体; 用户标签

中图分类号: TM71

文献标志码: A

文章编号: 1671-6841(2016)03-0113-05

DOI: 10.13705/j.issn.1671-6841.2016090

0 引言

在能源互联网和智能电网背景下^[1-3],准确地对用户用电行为特征分析和用电量短期、长期预测对电网需求侧管理和基于互联网的售电体精准营销具有重要的指导意义^[4-5].基于用户社会属性、用电行为特征、互联网行为将电力用户进行分类,进而实现群体用电行为预测,可以帮助售电体更深入地了解用户群体特征,为电力精准营销和制定电力需求侧响应策略提供数据支撑.

随着我国电力体制改革和能源互联网战略的不断推进,电力用户在互联网售电平台上购电和用电增值服务,形成种类丰富的用户数据^[6],包括用户属性数据、用电行为数据、上网行为数据等.基于这些用电大数据将用户用电行为特性进行多维度分解,建立用户分类模型,深刻认识电力用户群体效应,能够形成基于群体的用电模式差异化预测模型.

目前,对电力用户群体分析方法多是基于历史负荷数据,建立基于人工智能方法的负荷识别和预测模型^[7-9],再利用聚类的方法实现基于用电行为的用户群体分析^[10-11],或是基于行业或领域等用户属性将电力用户进行细分^[12-13],进而形成群体-属性的关系.这些分类方法往往一个群体中的用户只包含了个别共同属性,难以真正形成包含个体属性、用电行为和互联网行为等多种属性的相似用户群体,进而影响基于群体特征的用户用电行为分析预测模型的构建^[14-18].

标签作为一种用户行为的标识方法,蕴含了很多反映用户属性、用电偏好的信息^[19].本文通过对标签数据的分析,结合相同标签数量和负荷相似度形成用户关联度网络模型,得到相似电力用户群体,提取群体的主要特征以及发现群体中的重要用户,便于互联网售电体实施个性化增值服务产品和用电套餐推荐,进而提升电力服务质量.

1 电力用户群体识别与分析方法

1.1 电力用户行为关系模型

如图1所示,对于电力用户的个体行为刻画分为两类,一是电力用户的社会属性,包括家庭人口、住房

收稿日期:2016-06-15

基金项目:国家自然科学基金民航联合研究重点支持项目(U1333201),河南省基础与前沿技术研究计划项目(112300410090).

作者简介:许长清(1973—),男,河南郑州人,高级工程师,主要从事能源互联网技术研究,E-mail: zzplus@163.com;通讯作者:赵华东(1978—),男,河南开封人,副教授,主要从事智能工业技术研究,E-mail: huadong@zzu.edu.cn.

引用本文:许长清,赵华东,宋晓辉.基于大数据的电力用户群体识别与分析方法研究[J].郑州大学学报(理学版),2016,48(3):113-117.

面积、房屋均价、户内热源类型等,反映了电力用户对电力的消费能力和潜力,该类描述可以帮助售电体刻画用户的形象,进而对用户进行分类或归纳;二是电力用户的用电行为,包括历史负荷曲线、互联网购电行为、用户信用等级等,反映了电力用户的实际电能消费量、购电渠道等用电行为特征,动态展示了用户的用电行为特点,可以帮助售电体了解用户的用电模式和消费习惯.对于电力用户群体的分类和行为刻画,体现在基于用户个体行为,形成对群体典型负荷和群体标签的描述,其中群体标签包含了群体中用户共同的社会属性和用电行为.

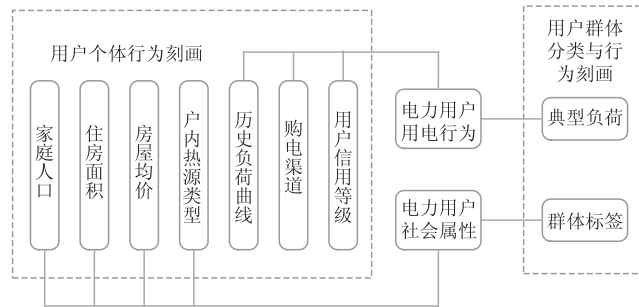


图1 电力用户行为关系模型
Fig.1 Model of power users' relation

1.2 相似用户群体识别方法

如图2所示,基于大数据的电力用户群体分析方法主要包括:

1) 构建电力用户及其标签二元关系网络.

标签是对用户属性、行为等一系列特征的一个或多个定性描述.一个标签通常是人为规定的高度精炼的特征标识,如电力用户的房屋面积标签:70~90 m²,人口标签:3人,信用等级标签:2级,等等.标签呈现出两个重要特征:语义化,能方便地理解每个标签含义,这也使得个体用户刻画模型具备实际意义;短文本,每个标签通常只表示一种含义,标签本身无需再做过多文本分析等预处理工作,这为利用机器提取标准化信息提供了便利.

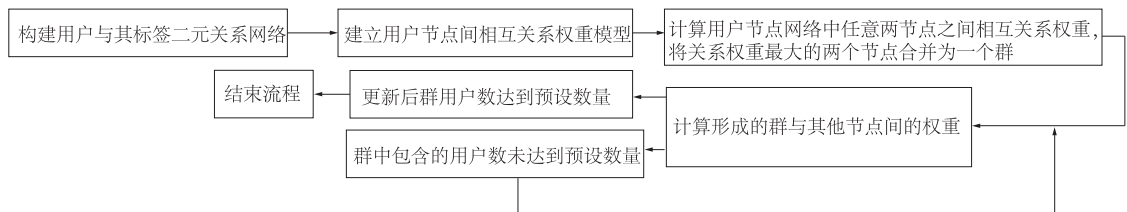


图2 电力用户群体分析方法
Fig.2 Method of power user group analysis

如图3所示,对电力用户全体分析的第一步是定义电力用户社会属性及用电行为标签,将用户与其所属的标签进行连接,形成相互关系二元网络,其中a、b、c、d代表用户节点,L1、L2、L3、L4代表用户标签.

2) 建立用户节点间相互关系权重模型.

用户节点间的相互关系权重是用户相似度的表现,包括图1所示的用户社会属性相似度和用电行为相似度,其中用户社会属性、用户信用和购电渠道等可以定义为静态标签,用相同标签数量表征其相似性,但用电行为中的历史负荷为动态数据,难以用标签描述,该指标却是用户相似性和群体刻画的关键内容.基于上述因素,本文以标签相同个数为权重模型的基础,将负荷相似度作为权重因子,使得权重模型既包含静态标签相似性,又包含动态负荷相似性.

将用户节点网络中任意两节点之间相同标签的个数表示为N_{ij},用电行为相似度S_{ij},两节点相互关系权重表示为

$$\alpha = N_{ij} \times S_{ij}, \tag{1}$$

其中:用电行为相似度S_{ij}的计算方法为:分别利用用户i和j日均负荷曲线中各小时的平均负荷组成用户日

负荷量 L_i 和 L_j , 基于余弦相似度计算方法计算用电行为相似度

$$S_{ij} = \frac{\sum_{n=1}^{24} (L_{in} \times L_{jn})}{\sqrt{\sum_{n=1}^{24} (L_{in})^2} \times \sqrt{\sum_{n=1}^{24} (L_{jn})^2}} \quad (2)$$

3) 识别与构建相似用户群体.

如图4所示,相似用户群体的形成过程包括:基于用户节点间相互关系权重模型,计算二元网络中任意两节点的权重,将权重最大的两个节点合并为一个群,并视其为网络中的一个节点,再与其他节点进行关系权重计算,将权重最大的节点合并至群中,形成新的用户群.如果群中的用户超过售电体设定的用户数,则完成相似用户群体的识别和构建,反之则继续计算更新的群与其他节点的关系权重,不断更新群组成.在计算群与节点关系权重过程中,标签只取群中用户共有特征,负荷量取群中所有用户的平均值,由于相似用户的识别过程是对网络中所有用户节点进行遍历计算,再进行权重排序,计算顺序将不影响模型结果.

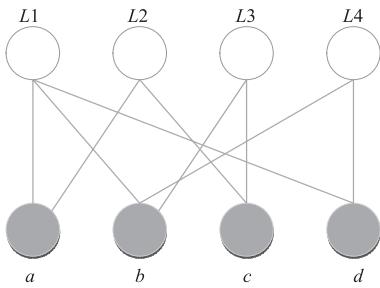


图3 “用户-标签”二元关系网络
Fig.3 “User-label” two-unit network

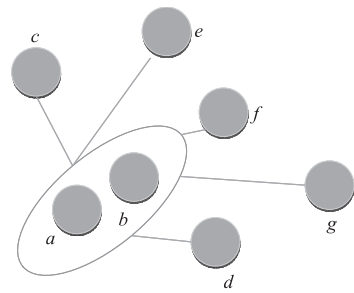


图4 相似用户群网络
Fig.4 Similar user group network

1.3 相似用户群体特征分析

在识别和构建相似用户群体后,一方面可以如图1所示,分析和获得群标签属性和用电行为特征,用以表征电力用户群体行为,实现售电体面向群体的销售决策;另一方面,相似用户群体的属性包含了个体用户的行为刻画,能够帮助售电体预测用户个体的用电行为,实现面向用户的电力精准营销.其中,群体标签为群用户共同拥有的标签类型,从群识别过程可以获得;群典型负荷量 L 反映了群体共同的用电行为特征,计算方法为

$$L = \sum_{i=1}^N \frac{L_i}{N} \quad (3)$$

其中: L_i 为相似用户群中用户 i 的日负荷量, N 为群中用户的数量.

通过群分析预测个体用户的用电行为,主要体现在对其负荷量的预测,对于新入网用户 j ,为了预测其用电行为特征,需要先识别该用户属于的群体,方法为分别计算用户 j 与所有群的相同标签个数,数值最大的群则为 j 的归属群,其负荷量

$$L_j = \sum_{i=1}^K \frac{L_i}{K} \quad (4)$$

其中: L_i 为所属群中与 j 相同标签个数最多的 K 个用户的负荷量, K 为售电体设定值,小于 N .

2 算例分析

本算例利用河南省某地区抽样 1 000 户居民的用电行为刻画数据,进行相似识别,分析群体特征,验证本文提出的方法.

2.1 电力用户标签体系和二元网络

如表1所示,用户标签体系主要包含家庭人口、住房面积、房屋均价、热源类型、用户信用、购电渠道等内

容.如图2所示,为每个用户赋予相应的标签,形成“用户-标签”二元网络,构建大数据分析系统.

2.2 相似用户群体识别与分析

利用1.2所述的相似用户群体分析方法,设定群用户数量 $N = 50$,识别出相似用户群体的共同标签特征包括:2~3人、60~90 m²、7 000~12 000 元/m²、燃气、市政热力.利用式(3)计算出群体的典型日负荷量,与50户个体负荷量各个小时的均方根误差均值见图5,各时刻平均为3.2%,表明群体典型日负荷量与个体负荷量误差较小,较为准确地体现了群体中个体普遍的用电行为特征.本文提出的群体识别方法能够有效识别出具有相似用电行为的用户群体.

表1 电力用户标签体系
Tab.1 Power user label system

用户行为属性	家庭人口/人	住房面积/m ²	房屋均价/元·m ⁻²	热源类型	购电渠道	用户信用
标签定义	1	<60	<7 000	燃气、市政热力 电力	营业厅 互联网	一级
	2~3	60~90 90~120	7 000~12 000			二级
	>3	>120	>12 000			三级

为了进一步验证通过对相似用户群体的分析可以用于预测新入网用户的用电行为特征,随机取实验样本以外的一个用户样本,提取样本用户标签后,寻找与该用户相同标签数量最多的用户群作为该用户的归属群,再分别计算该用户与归属群中其他用户的相同标签个数,取数值最大的前10个用户,计算其平均日负荷量,作为该用户的负荷量预测值.如图6所示,预测负荷值的曲线与实际值的曲线趋势相似,其均方根误差平均值为5.3%.结果表明,基于相似群体分类结果,利用新入网用户的属性标签,可以有效预测用户的用电行为趋势特征,对于售电体以用户为中心实施基于互联网的精准营销、制定需求响应策略、优化用户能效等具有支撑作用.

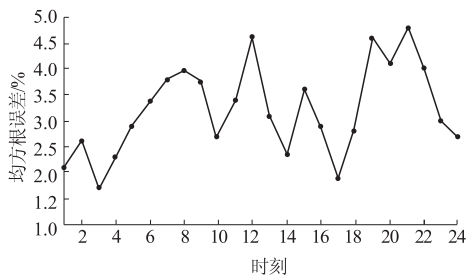


图5 群体典型负荷与个体负荷均方根误差
Fig.5 Mean square error curve of group typical load and individual load

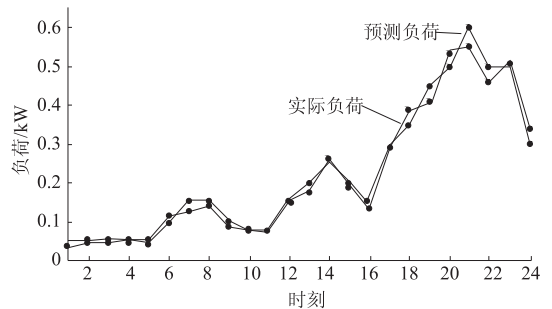


图6 负荷预测对比图
Fig.6 Load forecasting contrast curve

3 结论

本文研究了电力用户相似群体识别与特征分析方法.针对电力用户用电行为和社会属性特征,建立用户标签体系及“用户-标签”二元网络,并以标签对比结果为基础,将用户历史负荷相似度作为动态系数修成标签对比结果,构建用户相互关系权重模型,作为衡量用户相似度识别的标准,在二元网络中聚合用户形成相似群体,进而获得群属性标签和典型负荷特征,最后以相似群体为基础,基于新入网用户属性标签将其归入相似群,并进一步预测新用户的用电行为特征.仿真算例表明,该方法通过对标签数据的分析,可以得到相似用户群体,准确提取相似用户群用电行为特征以及发现群体中的重要用户,并预测新入网用户的负荷曲线趋势,便于售电体为电力用户提供个性化用电服务、精准实施需求响应策略、提升电网能效.

参考文献:

- [1] 严太山,程浩忠,曾平良,等. 能源互联网体系架构及关键技术[J]. 电网技术,2016,40(1):105-113.
- [2] 周海明,刘广一,刘超群. 能源互联网技术框架研究[J]. 中国电力,2014,47(11):140-144.
- [3] 曾鸣,杨雍琦,刘敦楠,等. 能源互联网“源-网-荷-储”协调优化运营模式及关键技术[J]. 电网技术,2016,40(1):114-124.
- [4] 何哲,王瑞峰,郝溪. 含分布式电源配电网单相接地故障测距研究[J]. 郑州大学学报(理学版),2016,48(1):85-90.
- [5] 马瑞,周谢,彭舟,等. 考虑气温因素的负荷特性统计指标关联特征数据挖掘[J]. 中国电机工程学报,2015,35(1):43-51.
- [6] 王继业,季知祥,史梦洁,等. 智能配用电大数据需求分析与应用研究[J]. 中国电机工程学报,2015,35(8):1829-1836.
- [7] CHICCO G, NAPOLI R, PIGLIONE F. Comparisons among clustering techniques for electricity customer classification[J]. IEEE Trans on Power Systems, 2006, 21(2):933-940.
- [8] 王德文,孙志伟. 电力用户侧大数据分析并行负荷预测[J]. 中国电机工程学报,2015,35(3):527-537.
- [9] 薛承荣,顾洁,赵建平,等. 基于用户用电特性及供电成本分摊的销售侧电价机制研究[J]. 华东电力,2014,42(1):168-173.
- [10] 黎祚,周步祥,林楠. 基于模糊聚类与改进BP算法的日负荷特性曲线分类与短期负荷预测[J]. 电力系统保护与控制,2012,40(3):56-60.
- [11] 彭夸,杨超. 基于用户实际日负荷曲线的负荷分类策略研究[J]. 贵州电力技术,2014,(2):48-49.
- [12] 蔡政权,刘至锋,管霖,等. 基于可靠性需求的电力客户细分和可靠性价值评估方法[J]. 广东电力,2015,28(5):44-50.
- [13] 张晓春,倪红芳,李娜. 基于数据挖掘的供电企业客户细分方法及模型研究[J]. 科技与管理,2013,15(6):104-109.
- [14] 廖志芳,王超群,李小庆,等. 张量分解的标签推荐及新用户标签推荐算法[J]. 小型微型计算机系统,2013,34(11):2472-2476.
- [15] 罗勇,李芳. 基于负荷预测的电力市场寡头竞争分析[J]. 郑州大学学报(理学版),2013,45(3):110-114.
- [16] 黄硕,楼树美. 基于XML路径表达式优化及其查询和过滤计算方法研究[J]. 信阳师范学院学报(自然科学版),2015,28(4):608-611.
- [17] 张晓煜,林晓,王志杰. 面向大数据正则表达式查询的有效算法[J]. 河南科技大学学报(自然科学版),2015,36(4):56-61.
- [18] 张素香,刘建明,赵丙镇,等. 基于云计算的居民用电行为分析模型研究[J]. 电网技术,2013,37(6):1542-1546.
- [19] 王海雷,俞学宁. 基于随机游走算法的社会化标签的用户推荐[J]. 计算机工程与设计,2013,34(7):2388-2391.

Research on Method of Power User Group Identification and Analysis Based on Large Data

XU Changqing¹, ZHAO Huadong², SONG Xiaohui³

(1. Economic Research Institute, Henan Electric Power Company, Zhengzhou 450002, China;

2. School of Mechanical Engineering, Zhengzhou University, Zhengzhou 450001, China;

3. Henan Academy of Sciences Institute of Applied Physics, Zhengzhou 450008, China)

Abstract: With the rapid development of Energy Internet and smart grid, it is an important part of “Internet plus” smart energy to have a good understanding of the power users group behaviors and providing accurate power service. Based on large data analysis of the power users, the user similarity weight model was built and the “user-label” two-unit network was put forward to identify similar groups. In addition, the group attributes and typical electricity consumption characteristics were obtained. This method could help to identify the important user groups, who might need customized electricity services and value-added services, for power suppliers. So the quality of power service could be greatly enhanced.

Key words: big data; power users; similar group; user label

(责任编辑:王浩毅)