

大数据环境下基于图嵌入的电力用户行为模式挖掘与预测

陈佐虎¹ 彭振国¹ 何海燕¹ 窦蕴韬¹
CHEN Zuohu PENG Zhengguo HE Haiyan DOU Yuntao

摘要

电力用户的行为具有极为显著的多样性与复杂性,传统的统计分析手段通常难以完整且精准地把握这些行为蕴含的内在规律。因此,文章提出大数据环境下基于图嵌入的电力用户行为模式挖掘与预测方法。首先,在大数据环境下收集并预处理电力用户的用电负荷数据;然后,根据预处理后的数据构建一个图结构模型,通过模型来挖掘电力用户的行为模式;最后,利用图嵌入算法从图结构模型中提取用户行为的特征,将这些特征输入到预测模型中,从而实现对电力用户行为模式的预测。实验结果表明,所设计的方法可以准确地挖掘出电力用户的用电行为模式,并且能够精准地预测出在不同电力用户行为模式下的用电负荷数据。通过这种方式,可以更好地理解电力用户的用电行为,并为电力系统的优化和管理提供有力的支持。此外,这种方法还具有很强的扩展性和适应性,能够应对未来电力系统中可能出现的各种复杂情况。

关键词

大数据环境;图嵌入;电力用户;行为模式挖掘;行为预测

doi: 10.3969/j.issn.1672-9528.2024.12.029

0 引言

随着智能电网技术的迅猛发展,电力系统正在经历一场深刻的变革,逐步向高度信息化和智能化的方向迈进。在这个过程中,电力系统积累了大量的用户用电负荷数据,这些数据不仅详细记录了用户的用电情况,还隐含了用户的消费习惯、生活模式,甚至反映出社会经济活动的规律,对于优化电力资源配置、提升电网运营效率以及实现个性化服务具有极其重要的意义。因此,如何有效地挖掘和利用这些数据,已经成为电力行业面临的一个重大挑战。在现有的研究中,文献[1]提出了一种基于社交网络用户行为数据构建决策树的方法,通过挖掘用户行为模式,取得了较高的精准性。然而,这种方法的可解释性较差,导致在实际应用中难以保障数据挖掘的效果。而文献[2]则将数据分析和机器学习算法结合起来,进行用户行为模式的挖掘,虽然准确性较高,但这种方法往往需要大量的计算资源。在实际应用中,由于资源受限,可能会导致挖掘结果出现较大的误差。针对上述研究中存在的局限性,本文提出了一种在大数据环境下基于图嵌入的电力用户行为模式挖掘与预测方法。这种方法不仅能够有效挖掘用户行为模式,还能够预测未来的用电行为,为电力企业的决策提供更加精准和可靠的数据支持。通过这种方法,电力企业可以更好地理解用户需求,优化电力资源配置,提

升电网运营效率,实现更加个性化的服务,从而在激烈的市场竞争中占据有利地位。

1 大数据环境下收集并预处理用电负荷数据

在大数据环境下,高级计量基础设施(AMI)系统是智能电网的关键技术之一,可以对电力用户的用电负荷数据进行实时监测与采集,所以本文在进行电力用户行为模式挖掘与预测时,首要任务是通过AMI系统采集电力用户的用电负荷数据,该系统通过智能电表等终端设备,收集包括但不限于用户编号(作为数据的主键标识)、时间戳(精确记录数据生成的时间点)、电量消耗(直接反映用户用电行为的核心指标)等基本信息。但是由于大数据环境下,原始收集的电力用户用电负荷数据规模极大,所以需要进行数据清洗。这里本文采用了IQR(四分位距)方法来识别并修正数据中的错误,假设原始电力用户用电负荷数据的第一四分位数为 q_1 ,第三四分位数为 q_3 ,那么数据的四分位距为:

$$q = q_3 - q_1 \quad (1)$$

式中: q 表示原始电力用户用电负荷数据的四分位距。根据式(1)确定数据的四分位距后,设定 $q \times k$ 为异常范围,其中 k 为常数,通常取值为1.5和3,将超过该范围的数据视为异常的错误值,并进行剔除即可。在完成海量电力用户用电负荷数据的异常识别与修正之后,由于不同用户之间的用电习惯、不同时间段内的负荷特性均存在显著差异,这些差异在数据挖掘过程中可能导致模型性能下降,所以本文还需对

电力用户用电负荷数据进行归一化处理^[3]，表达式为：

$$X = \frac{X' - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

式中： X' 、 X 分别表示处理前、后的电力用户用电负荷数据； $X_{\min}$ 、 $X_{\max}$ 分别表示原始电力用户用电负荷数据的最小和最大值。经过上述一系列收集与预处理步骤后，本文可以得到一组高质量且标准化的电力用户用电负荷数据集，为后续用户行为模式挖掘与预测提供基础。

2 构建图结构模型挖掘电力用户行为模式

基于电力用户用电负荷数据，本文设计了一种基于图结构模型的电力用户行为模式挖掘方法^[4]，该模型不仅有助于分析用户行为的内在规律，还可以为后续的用户行为预测提供坚实基础。首先，构建一个图结构模型，将每个电力用户视为图中的一个节点，表示为集合 $D = \{d_1, d_2, \dots, d_n\}$ ，其中 n 是用户总数。同时，用户间的相似性或交互关系通过边来表示，边的权重 ω_{ij} 反映了用户 d_i 和 d_j 之间关系的紧密程度，计算公式为：

$$\omega_{ij} = \alpha \cdot \sin_{\text{geo}}(d_i, d_j) + \beta \cdot \sin_{\text{elec}}(d_i, d_j) \quad (3)$$

式中： α 、 β 表示权重系数； $\sin_{\text{geo}}(d_i, d_j)$ 、 $\sin_{\text{elec}}(d_i, d_j)$ 分别表示电力用户的地理位置和用电习惯的相似度函数。根据边和权重可以构建图结构的邻接矩阵 W ，其中 $W_{ij} = \omega_{ij}$ ，若 $i = j$ 则 $W_{ij} = 0$ （无自环）。在根据上述内容构建出图结构模型后，本文采用谱聚类算法进行用户行为模式的挖掘：根据图结构的邻接矩阵 W ，可以进一步确定节点的度矩阵 Z ，表达式为：

$$Z_{ii} = \sum_{j=1}^n W_{ij} \quad (4)$$

式中： Z_{ii} 表示电力用户节点 d_i 的度。在式（4）所求度矩阵基础上，可以构建一个拉普拉斯矩阵 P ，其公式为：

$$P = Z - W \quad (5)$$

然后，对式（5）所求拉普拉斯矩阵进行特征分解，得到特征值和特征向量，选择前 k 个最小非零特征值对应的特征向量（这里 k 是目标聚类数），构成新的特征空间。最后，在新的特征空间中对各个电力用户节点进行聚类，得到 k 个聚类簇，观察聚类簇内节点间的连接强度、聚类结构等特征，即可了解不同簇内用户的行为模式，以此实现电力用户行为模式的挖掘。

3 基于图嵌入预测电力用户行为模式

根据上述内容挖掘出电力用户行为模式后，为了降低数据维度并减少计算复杂度，本文在电力用户行为数据的图结构模型基础上，采用图嵌入技术^[5]，将高维的图结构信息映射到低维的向量空间中，同时能够保留保留节点间的相似性和图的整体结构信息。这里引入了图嵌入中的常用算

法-Node2Vec算法，通过模拟节点在图上的随机游走过程，学习节点的低维向量表示。在利用Node2Vec算法提取电力用户行为数据特征时，通常定义最大化给定节点序列中相邻节点的条件概率作为目标函数，其公式为：

$$\max_{\varphi} \sum_{(d_i, d_j) \in D} \log f(d_i | d_j; \varphi) \quad (6)$$

其中：

$$f(d_i | d_j; \varphi) = \frac{\exp(S(d_i, d_j))}{\sum_{d_s \in D} \exp(S(d_i, d_s))} \quad (7)$$

式中： $f(d_i | d_j; \varphi)$ 表示在给定节点 d_j 的条件下，通过节点嵌入向量参数 φ 预测节点 d_i 出现的概率； $S(d_i, d_j)$ 、 $S(d_i, d_s)$ 分别表示节点 d_i 和 d_j 、节点 d_i 和 d_s 嵌入向量的相似度量。根据上式即可将原始电力用户行为数据图结构中的节点映射到低维向量空间中，提取到用户行为特征^[6]。获得图嵌入后的用户节点向量后，本文进一步利用这些向量作为特征输入到预测模型中，以预测用户未来的行为模式。这里本文将电力用户行为模式预测模型表示为 F ，给定电力用户行为模式 $Y = \{y_1, y_2, \dots, y_m\}$ ，假设提取的电力用户用电行为特征集合为 $X = \{x_1, x_2, \dots, x_n\}$ ，对于任意特征 x_i ，将其输入到模型 F 中，模型对应的输出 G_{x_i} 为特征 x_i 属于行为模式 y_j 的概率，其公式为：

$$G_{x_i} = F(x_i) \quad (8)$$

对于电力用户用电行为特征 $X = \{x_1, x_2, \dots, x_n\}$ ，该预测模型可以一一预测并输出各特征属于各个电力用户行为模式 $Y = \{y_1, y_2, \dots, y_m\}$ 的概率分布，取最大概率值所对应的模式作为最终预测结果，以此完成电力用户行为模式预测。

4 仿真实验

4.1 实验准备

为验证大数据环境下基于图嵌入的电力用户行为模式挖掘与预测方法的有效性和正确性，本章引入基于决策树的电力用户行为模式挖掘与预测方法和基于机器学习的电力用户行为模式挖掘与预测方法，作为对照组，展开一个仿真对比实验。首先，搭建本次实验的仿真环境，具体参数配置如表1所示。

表1 实验环境配置表

实验环境	参数配置
硬件环境	Intel Core i7-10700K CPU、32 GB RAM、NVIDIA GeForce RTX 3080 GPU
软件环境	Python 3.8、TensorFlow 2.5、scikit-learn 0.24、Pandas 1.2.4、NumPy 1.19.5

然后，采用某电力公司提供的真实负荷数据作为实验数

据, 包含了 1485 个电力用户用电负荷样本数据。数据集被划分为 6 个不同的电力用户行为模式, 每个模式下的训练样本和测试样本数量分布如表 2 所示。

表 2 实验数据分布表

标签	行为模式	训练样本数	测试样本数
1	居民用电模式	160	40
2	办公用电模式	180	45
3	商业用电模式	200	50
4	工业用电模式	200	55
5	医院用电模式	210	60
6	农业用电模式	220	65

最后, 分别将本文设计方法和对照组中两种方法应用于实验数据集上, 展开电力用户行为模式的挖掘与预测, 对比分析实验结果。

4.2 结果分析

在本次电力用户行为模式挖掘与预测实验中, 为直观且客观地评价实验组方法和对照组中两种方法的性能, 本文分别引入准确率和决定系数作为电力用户行为模式挖掘结果和预测结果的评价指标, 具体计算公式为:

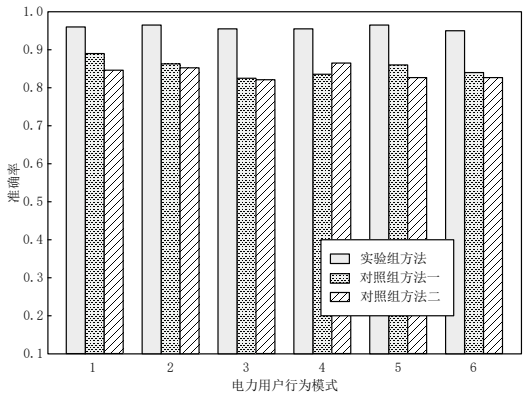
$$A = \frac{1}{N_0} \sum_{i=1}^{N_0} \gamma_i 1(y'_i = y_i) \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N_0} (Y'_i - Y_i)^2}{\sum_{i=1}^{N_0} (\bar{Y} - Y_i)^2} \quad (10)$$

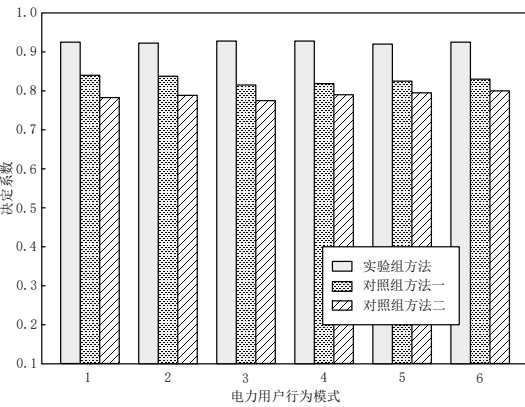
式中: A 表示电力用户行为模式挖掘结果的准确率; N_0 表示电力用户数量; γ_i 表示第 i 个电力用户的行为模式在日负荷曲线中的概率分布; y'_i 、 y_i 分别表示第 i 个电力用户的行为模式挖掘值和实际值; R^2 表示电力用户行为模式挖掘结果的决定系数, 取值范围为 $[0,1]$, 其值越大表明预测结果和实际数据拟合效果越好, 也就是预测精度越高; Y'_i 、 Y_i 分别表示第 i 个电力用户的行为模式预测值和实际值; $\bar{Y}$ 表示电力用户行为模式的平均值。

因此, 在完成电力用户行为模式挖掘与预测后, 采用上式分别计算实验组方法和对照组中两种方法所得挖掘结果的准确率、预测结果的决定系数, 作为本次实验结果, 如图 1 所示。从图 1 中可以看出, 无论是在电力用户行为模式的挖掘中, 还是在电力用户行为模式的预测中, 本文设计方法均表现最佳。具体而言, 本文设计方法下, 电力用户行为模式挖掘结果的准确率达 0.959, 较对照组中两种方法分别提升了 0.108、0.120; 同时, 电力用户行为模式预测结果的决定系数高达 0.925, 较对照组中两种方法分别提

升了 0.097、0.137。因此, 本文研究的大数据环境下基于图嵌入的电力用户行为模式挖掘与预测方法是可行且可靠的。



(a) 电力用户行为模式挖掘结果



(b) 电力用户行为模式预测结果

图 1 电力用户行为模式挖掘与预测结果对比

5 结语

在本次研究中, 深入探讨了在大数据环境下, 基于图嵌入技术的电力用户行为模式挖掘与预测方法。即通过收集并预处理电力用户用电负荷数据, 确保数据的准确性和可用性。接着, 结合数据挖掘技术和图嵌入技术, 成功实现了对电力用户行为模式的深入挖掘与准确预测。

在未来的研究中, 将致力于进一步优化图嵌入算法, 以提高其对复杂用户行为模式的捕捉能力。将重点关注算法的效率和准确性, 力求在处理大规模数据时仍能保持高性能。此外, 还将探索将更多外部因素(如社会经济状况、气候变化等)融入预测模型, 以增强模型的实际应用价值。通过引入这些外部因素, 模型将能更全面地理解用户行为, 从而提供更准确的预测结果。此外, 本文还将关注电力用户隐私保护问题。在确保数据安全的前提下, 将采取各种措施来保护用户隐私, 例如使用匿名化技术、加密技术等。在保护用户隐私的同时, 推动研究成果在智能电网领域的广泛应用, 为智能电网的发展做出贡献, 提高电力系统的效率和可靠性。

融合 GRU-TCN 的深度知识追踪方法

谭 雪¹

TAN Xue

摘 要

知识追踪 (KT) 作为智能教育系统中的核心技术, 致力于通过分析学生的历史答题记录, 动态建模其对特定知识点的理解与掌握情况。然而, 当前大多数知识追踪模型在应对长序列数据和复杂时间依赖性时仍存在显著的局限性。为此, 文章提出了一种融合时域卷积网络 (TCN) 与门控循环单元 (GRU) 的全新知识追踪模型 TGC-KT。TCN 模块通过捕捉局部时间依赖性, 显著提升了模型在处理长期依赖关系时的能力, 而 GRU 则凭借其门控机制对历史信息进行选择记忆与遗忘, 进一步优化了预测精度。此外, TGC-KT 能够有效结合局部与全局的学习信息, 动态调整模型对学生知识状态的估计。实验结果表明, 在两个公开数据集上的测试中, TGC-KT 模型在知识掌握度评估和学习效果预测方面的表现优于现有方法, 展示了其在个性化教育中的潜在应用价值。

关键词

深度学习; 知识追踪; 预测精度; 教育数据挖掘

doi: 10.3969/j.issn.1672-9528.2024.12.030

0 引言

随着个性化学习需求的日益增长, 知识追踪 (KT) 技术在教育领域的重要性愈发凸显^[1]。KT 通过分析学生的动态知识变化, 为学生提供个性化学习路径, 同时帮助教师优化教学策略。近年来, 全球范围内的新冠肺炎 (COVID-19) 疫

情对教育系统产生了深远影响, 迫使学校和教育机构迅速转向远程教学模式。这一转变不仅对教育资源的分配提出了新的要求, 也为 KT 技术的发展带来了新的机遇和挑战。

KT 旨在通过建模学生的历史序列数据来得到学生当前的知识状态。利用学生与问题之间的交互信息, 来预测学生是否能正确回答下一个相关的问题。如图 1 所示, 学生回答问题可能包含多个技能知识点, 若学生在做题过程中能将已经获得的先验知识点串联起来回答问题, 则学生大概率能答对这个问题。这种评价方式可以帮助教师了解学生的学习情况, 对学生答题错误的原因进行分析, 帮助教师为学生制定

1. 贵州民族大学 贵州贵阳 550000

[基金项目] 贵州省教育厅 (黔教技〔2022〕047 号), 贵州省高等学校智慧教育工程研究中心; 贵州民族大学自然科学基金 (GZMUZK〔2021〕YB22)

参考文献:

- [1] 韩永印, 王侠, 王志晓. 基于决策树的社交网络隐式用户行为数据挖掘方法 [J]. 沈阳工业大学学报, 2024, 46(3): 312-317.
- [2] 吴自博, 王波, 陈清, 等. 基于机器学习的调度操作行为挖掘与推荐技术研究及应用 [J]. 电力系统自动化, 2022, 46(8): 181-188.
- [3] 严玉婷, 薛冰, 方力谦, 等. 基于周期调整负荷成分指数的行业用电大数据价值挖掘 [J]. 电力科学与技术学报, 2022, 37(6): 181-189.
- [4] 张敬伟, 刘绍建, 杨青, 等. DMFUCP: 大规模轨迹数据通用伴随模式分布式挖掘框架 [J]. 计算机研究与发展, 2022, 59(3): 647-660.
- [5] 梁瑶, 谢春丽, 王文捷. 基于图嵌入的代码相似性度量 [J]. 计算机科学, 2022, 49(S2): 801-806.

- [6] 侯靖琳, 仇润鹤, 薛季爱, 等. 基于知识图谱嵌入和补全的电梯故障预测 [J]. 计算机工程与设计, 2022, 43(1): 224-230.

【作者简介】

陈佐虎 (1989—), 男, 甘肃白银人, 本科, 工程师, 研究方向: 电网数字化。

彭振国 (1984—), 男, 甘肃正宁人, 本科, 工程师, 研究方向: 电网数字化。

何海燕 (1995—), 女, 甘肃定西人, 本科, 工程师, 研究方向: 电力大数据。

窦蕴韬 (1989—), 男, 甘肃兰州人, 本科, 工程师, 研究方向: 电网数字化。

(收稿日期: 2024-09-05)