

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

吴自博^{1,2}, 王波^{1,2,3}, 陈清^{1,2}, 郭耀松^{1,2}, 赵京虎^{1,2}, 闪鑫^{1,2,3}

(1. 南瑞集团有限公司(国网电力科学研究院有限公司), 江苏省南京市 211106; 2. 国电南瑞科技股份有限公司, 江苏省南京市 211106; 3. 智能电网保护和运行控制国家重点实验室, 江苏省南京市 211106)

摘要: 受互联网用户推荐技术的启发,基于调度系统中运行人员的历史操作记录,结合数据分析与机器学习算法提出了一种有效的调度操作行为模式挖掘与推荐技术。首先,所提技术采集记录调度员日常操作数据,并通过页面类型划分、会话识别、事务识别等过程实现数据预处理。然后,利用模式分析和聚类分析完成调度员行为模式挖掘,并结合关联规则将模式数据压缩至频繁模式树(FP-tree)。最后,利用基于模式树结构的实时推荐系统为调度员提供操作推荐服务。在实际调度系统上进行验证和应用,结果表明所提技术能有效识别出调度员的操作行为模式并为调度员提供较为精准的操作推荐,进而提高调度员与调度系统的交互效率,减少调度员的工作量。

关键词: 电网调度; 机器学习; 行为模式; 操作日志; 聚类挖掘; 操作推荐

0 引言

调度系统作为现代电网平稳运行的基石,涵盖了输、配电等调度业务和对电网整体运行状态的监测、控制、管理等任务。调度员通过调度系统监控电网运行状况,分析电网故障信息,保障电网的安全与稳定。但随着电网规模的不断扩大、调度系统的迭代升级、操作程序的日趋复杂,调度员的工作量和工作负担也逐渐加重,尤其是调度员在调控系统上进行断面监视、负荷控制、方式调整及故障分析等工作时,往往需要花费大量精力在多个系统页面之间进行反复跳转、查找,以获取某个业务场景所需指标,或在特定页面上实施重复烦琐的操作流程,严重影响了调度员的工作效率^[1-4]。因此,如何优化调度员的操作流程、简化调度员操作任务,对于减少调度员工作负担、提高调度工作效率具有重要意义,值得深入研究^[5]。

随着大数据的普及和发展,近年来基于用户日志的数据挖掘和推荐技术日趋成熟,并在许多领域取得了成功应用。文献[6-7]通过分析用户在网站上的浏览路径,对网站的页面结构进行优化,方便用户快速到达其感兴趣的页面。文献[8-9]通过对用户的浏览记录进行聚类分类,实现对不同类型用户推荐不同内容的功能。文献[10]根据期刊的推荐业务需求,设计了数字期刊推荐系统的框架,实现期刊

内容资源与用户需求的有效匹配。

在电网调度领域,目前鲜有对调度操作行为模式挖掘和推荐的技术研究,更多的是基于调度规程和人工经验的操作辅助决策。文献[11]指出了当前电网调度的一些业务模式及相应规范准则,但不涉及调度员操作模式挖掘及效率提升。文献[12]采用自然语言处理技术对线路跳闸的处理规程进行知识抽取,形成流程化的知识,用于协助调控人员完成故障处理。文献[13-14]提出通过构建调控领域的知识图谱,支撑调度员在故障处置、倒闸操作、对话问答等业务场景的辅助决策,有效地降低人工处置的风险。

因此,受互联网用户个性化推荐思路的启发,本文基于调度员的历史操作行为数据,采用机器学习技术与数据挖掘算法对调度员操作行为模式挖掘进行了深入研究,并在此基础上实现了基于操作模式挖掘结果的调度智能推荐,有效简化了调度员的操作步骤,节省了操作时间,提升了工作效率。

1 整体框架

基于机器学习的调度操作行为模式挖掘与操作推荐总体框架如图1所示,主要包括:数据采集、模式挖掘、实时推荐、交互界面4个模块。

1) 数据采集模块对不同用户在系统界面中的操作动作进行记录和管理。行为捕捉以内部埋点跟踪和事件触发的方式捕捉调度员的操作动作;行为处理将捕捉到的操作以行为日志和实时消息的形式生

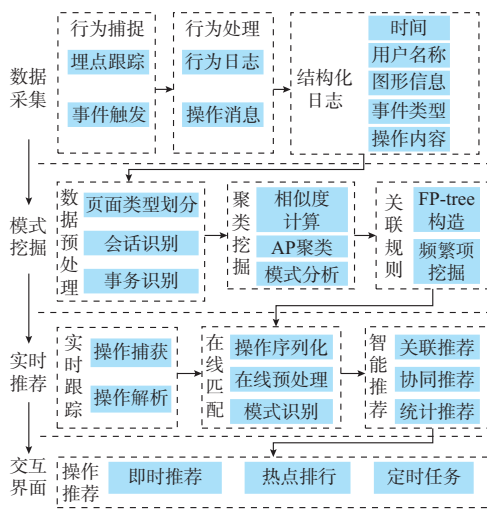


图1 电网调度行为模式挖掘及推荐框架
Fig. 1 Mining and recommendation framework for power grid dispatch behavior pattern

成相应结构化操作日志。如附录 A 图 A1 所示,操作日志通过预设标签标识用户名、操作时间、图形名称、操作内容等信息。

2) 模式挖掘模块负责对历史操作记录中的调度员浏览和操作模式进行挖掘。其中,数据预处理用以将原始的调度员操作记录处理成后续模型需要的用户会话;聚类挖掘对用户会话进行挖掘分析获得相应的操作行为模式;关联规则模块以聚类分析出的模式为基础,构造该模式的频繁模式树(FP-tree)并挖掘出频繁模式项作为后续的推荐数据源。

3) 实时推荐模块通过实时跟踪用户页面操作序列,在线匹配识别用户所处操作行为模式,完成对后续操作的关联推荐、统计推荐等智能操作推荐服务。

4) 交互界面负责操作推荐结果的实时展示与操作交互,提供即时推荐、热点排行、定时任务等多种操作推荐方式,并可通过快捷操作的交互方式实现对推荐内容的一键操作。

2 关键技术

行为模式挖掘与操作实时推荐是整个技术框架的核心内容,下面将介绍其涉及的关键技术。

2.1 数据预处理

2.1.1 页面类型划分

在电网调度系统中,有的类型页面可以展示不同场站、设备的内容。例如所有“XX 变电站 .fac”表示各具体厂站图,即厂站类型的页面。在操作模式挖掘中需将这样的页面进行归一化处理,避免将具体厂站、设备信息混杂在操作模式中,影响后续模式的聚类效果。

通常系统页面可分为应用系统图、厂站图、曲线

图等不同类型。页面类型划分时,对于逻辑上具有唯一性的应用系统图,采用不同页面编号来划分定义;而对于具有不同页面实例的厂站图(如 XX 变电站厂站图)、曲线图(如 XX 设备有功功率今日曲线)等页面类型,按照其所属页面类型定义为相同的页面编号。

2.1.2 会话识别

调度员的操作行为记录是时间上连续的一串页面浏览序列,会话为按照不同调度业务从整串页面浏览序列中截取出的子序列,调度员通过浏览子序列中的页面完成某个任务。会话中的单个页面及其停留时长构成相应的一个页面实例。

本文采用时间窗口法进行会话识别^[15]。时间窗口法认为两个连续页面之间的时间间隔大于某个阈值时,原会话截止,新的会话产生。在调度系统中,调度员连续操作页面进行跳转时,若某一页面与其下一页面的间隔时间大于设定阈值,则认为该页面与其下一页面分属不同的会话,通常会话识别默认阈值为 300 s^[16]。

在电网调度系统中,调控人员为了监视某些指标,经常在一些页面长时间停留,远超过 300 s,会话并未就此断开。针对该特点,本文提出对于不同页面通过分析其页面实例停留时长的分布区间给出不同的会话识别阈值。具体地,定义页面实例停留时长的中位数为页面参考时长,当页面的参考时长超过 300 s 时,则其会话识别阈值调整为默认阈值的 2 倍,即 600 s。

2.1.3 事务识别

在页面浏览序列中,根据用户浏览性质可将页面分为事务页和导航页。事务页是具有访问目的、调控人员感兴趣的页面,如电网负荷监视页;而导航页是调控人员在浏览过程中跳转性质的页面,如厂站目录页。所谓事务识别就是区分出页面浏览序列中的事务页。事务识别如图 2 所示,其中 N 表示导航页,C 表示事务页。

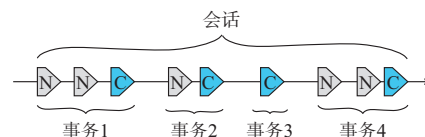


图2 事务识别示意图
Fig. 2 Schematic diagram of routine identification

页面实例的停留时长超过设定阈值时,该页面实例是事务页;反之该页面实例是导航页^[17]。其中,阈值的设定通过分析典型事务页和导航页的页面停留时长分布状况给出。其设定过程同会话识别

类似,根据二八原则,定义80%的页面实例不超过的停留时长为事务参考值。

2.2 聚类挖掘

经过上文数据预处理后,调度员的操作日志被分段逐条整理成经过编号的浏览序列,然后通过聚类算法将相似的浏览序列划分为一类,从而识别出调度员的浏览行为模式。本文采用近邻传播(AP)聚类算法进行模式挖掘^[18]。该算法无须预先指定聚类的簇数,输入参数为两两会话之间的相似度,以解决多个会话向量维度无法统一的问题。

2.2.1 会话相似度计算

会话相似度计算的本质方法为余弦相似度的计算,首先根据0-0匹配原则对2个会话建立非零的会话向量^[19],其过程如下。

建立2个事务会话集 a 、 b 及 a 、 b 的并集 c , $a = \{a_1, a_2, \dots, a_n\}$, $b = \{b_1, b_2, \dots, b_m\}$, $c = \{c_1, c_2, \dots, c_k\}$,其中, n 、 m 、 k 分别为 a 、 b 、 c 中事务的个数。基于并集 c 可以将集合 a 、 b 转换为向量 x 、 y 的形式表示,即 $x = [x_1, x_2, \dots, x_k]$, $y = [y_1, y_2, \dots, y_k]$,其向量元素 x_k 和 y_k 分别表示会话集 a 、 b 中第 k 个事务在事务并集 c 中出现的次数。

考虑到存在用户需要反复在几个事务页间切换的情况,事务页出现次数过高,会使得相似会话的相似度降低,影响分类结果。针对该问题,本文提出对会话向量中的元素进行频次加权处理:

$$\bar{x}_k = (x_k - 1)w + 1 \quad (1)$$

$$\bar{y}_k = (y_k - 1)w + 1 \quad (2)$$

式中: \bar{x}_k 和 \bar{y}_k 分别为 x_k 和 y_k 的修正值; w 为权值,可通过分析案例会话中的事务重复程度来设置,重复度越高,权值越小,本文根据事务重复度设置权值为0.1。

然后,计算会话相似度为:

$$s(x, y) = \frac{\bar{x}_1 \bar{y}_1 + \bar{x}_2 \bar{y}_2 + \dots + \bar{x}_k \bar{y}_k}{\sqrt{\bar{x}_1^2 + \bar{x}_2^2 + \dots + \bar{x}_k^2} \sqrt{\bar{y}_1^2 + \bar{y}_2^2 + \dots + \bar{y}_k^2}} \quad (3)$$

2.2.2 AP聚类算法

AP聚类算法是基于数据点间的信息传递的一种聚类算法^[20],主要通过迭代更新吸引度矩阵和归属感矩阵进行。其中,吸引度矩阵描述数据对象适合作为聚类中心的程度;归属感矩阵描述数据对象选择其他对象作为其聚类中心的适合程度。

吸引度矩阵中元素 $r_{t+1}(i, k)$ 按照式(4)迭代:

$$r_{t+1}(i, k) = s(i, k) - \max_{k' \neq k} (a_t(i, k') + s(i, k')) \quad (4)$$

归属感矩阵中元素 $a_{t+1}(i, k)$ 按照式(5)迭代:

$$a_{t+1}(i, k) = \min \left(0, r_t(k, k) + \sum_{i' \in \{i, k\}} \max(0, r_t(i', k)) \right) \quad (5)$$

式中: $r_{t+1}(i, k)$ 表示数据对象 k 适合作为数据对象 i 的聚类中心的程度; k' 为除 k 外的其他数据对象; $a_{t+1}(i, k)$ 表示数据对象 i 选择数据对象 k 作为其聚类中心的合适程度; i' 为除 i 外的其他数据对象; $s(i, k)$ 为数据对象 i 和 k 的相似度。

为了避免振荡,AP算法在迭代更新过程中引入了衰减系数 λ :

$$r_{t+1}(i, k) \leftarrow (1 - \lambda)r_{t+1}(i, k) + \lambda r_t(i, k) \quad (6)$$

$$a_{t+1}(i, k) \leftarrow (1 - \lambda)a_{t+1}(i, k) + \lambda a_t(i, k) \quad (7)$$

如果经过若干次迭代之后保持不变或者算法执行超过设定的迭代次数,则算法结束。

聚类效果以轮廓系数为评价标准^[21]。对于个体点其轮廓系数 S_i 为:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (8)$$

式中: b_i 为第 i 个对象和不包含该对象的任意簇中所有对象的平均距离的最小值; a_i 为第 i 个对象到所属簇中所有其他对象的平均距离。

2.3 关联规则

实时推荐时需要用户对用户操作序列和挖掘出的行为模式在线逐个匹配并计算置信度,因此,选取一种合适的数据结构缓存行为模式的相关信息对于提升匹配和推荐效率至关重要。本文采用频繁模式增长(FP-Growth)算法来建立FP-tree数据缓存结构并挖掘关联规则^[22]。

FP-Growth算法通过对源数据进行2次扫描,将数据信息存储至FP-tree,从而避免产生候选项集的步骤,减少了数据匹配开销。该算法由构建FP-tree和频繁项集挖掘2个步骤组成。FP-tree是一种扩展的前缀树结构,由频繁项头指针表和项前缀树构成,存储关于频繁模式次数的信息;频繁项集挖掘是通过头指针表生成条件模式基后,根据计数值判定该路径集合是否为频繁项,迭代重复上述步骤,递归地发现频繁项、条件模式基以及另外的条件树。其中,挖掘出的频繁项集是用户日常工作中的页面关联规则。在电网调控领域,用户的页面浏览顺序和业务背景有密切关系,但该频繁项集无法反应页面的浏览顺序。因此,本文参考历史会话中的页面浏览顺序对关联规则中的页面进行重新排序,筛选出包含关联规则中所有页面的最新历史会话,以最新历史会话中的页面浏览顺序作为该关联规则的页

面顺序。

2.4 实时推荐

实时推荐以挖掘出的调度员行为模式为基础,结合调度员最近的操作行为序列推荐出后续操作,包括模式识别、关联推荐等,整体流程如图3所示。

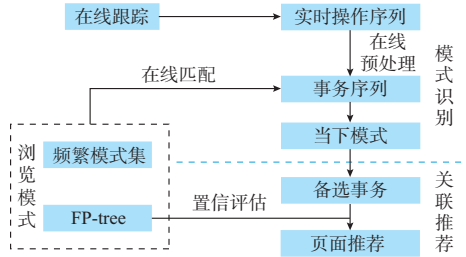


图3 实时推荐流程
Fig. 3 Process of real-time recommendation

2.4.1 模式识别

模式识别用于判断当前用户所处的操作模式。首先,通过在线跟踪捕获用户在系统页面上的操作,形成实时操作序列;然后,通过在线预处理获得调度员当前正在进行的会话事务序列;最后,遍历各行为模式的频繁项集,进行模式匹配。若实时会话序列和判定行为模式的频繁项存在交集,则该行为模式可能为当前实时会话序列的匹配模式,选取其中匹配度最高的若干行为模式作为模式识别的结果。

2.4.2 关联推荐

关联推荐以模式识别生成的当下模式和FP-tree为基础进行,分析出当前用户行为序列置信度最高的操作事务。首先,通过比对模式识别结果,生成备选事务;然后,对备选事务进行置信评估,获得发生概率最高的若干备选事务。具体地,搜索FP-tree逐个计算备选推荐事务的置信度,选取置信度前N名的事务项作为页面推荐内容。

3 验证与应用

本文采用某电力调度中心智能电网调度技术支持系统(D5000)中的调度员历史操作数据进行推荐应用与算法验证。

3.1 算法验证

3.1.1 页面浏览统计

本节以浏览行为模式挖掘为例介绍算法验证情况。收集的数据时间范围为2020年6月1日至6月30日。该时间段内共采集操作记录5 932条,涉及页面241个。对上述241个页面进行浏览次数统计,排名前十的页面如图4所示。从图中可以看出,调度员浏览次数最多的为“电网负荷监视.fac”、“自动发电控制(AGC)-运行监控.agc”、“系统首页.sys”等页面。

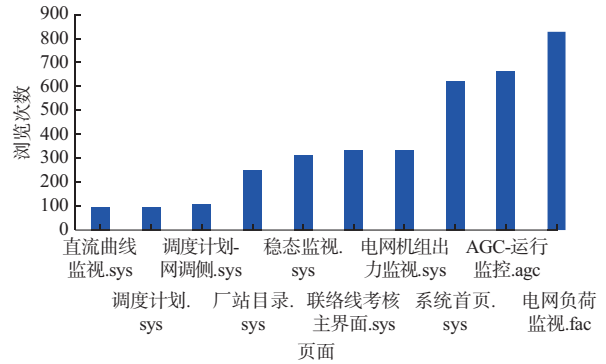


图4 页面浏览次数排名
Fig. 4 Pageview ranking

3.1.2 页面类型划分

以点击次数前10的页面为例,其页面类型划分结果如表1所示。其中,类型对应该页面在后续模式挖掘中的编号。

表1 页面类型划分示例
Table 1 Example of page type division

页面名称	类型
直流曲线监视.sys	12
调度计划.sys	7
调度计划-网调侧.sys	6
厂站目录.sys	3
稳态监视.sys	2
联络线考核主界面.sys	4
电网机组出力监视.sys	5
系统首页.sys	1
AGC-运行监控.agc	531
电网负荷监视.fac	494

3.1.3 会话识别结果

对研究样本数据通过前述会话识别、事务识别、页面类型划分后,梳理出有效会话196个。以2020年6月1日为例,选取发生的典型会话,如表2所示。其中,事务页序列为会话中事务页的页面类型划分编号。

表2 会话识别结果
Table 2 Session recognition results

会话编号	发生时间	事务页序列
9448	2020-06-01 00:02	5,494,4,494,5,6
9450	2020-06-01 07:47	531,494,683,494,683
9465	2020-06-01 10:22	34,531,34,531,12
9466	2020-06-01 13:15	5,494,5,494,4,494,5,314
9468	2020-06-01 15:09	531,4,494,531,626,4,15,626
9469	2020-06-01 16:07	494,8,34,494,683,8
9474	2020-06-01 18:51	494,9,494,9,18
9478	2020-06-01 21:02	4,5,626,5,12

3.1.4 会话相似度计算

以表2中的9448、9450、9465、9466会话为例,计算对应相似度矩阵如表3所示。矩阵中数字为该数值对应的行和列上的会话的相似度。

表3 相似度矩阵
Table 3 Similarity matrix

会话	相似度			
	9448	9450	9465	9466
9448	1.00	0.31	0	0.88
9450	0.31	1.00	0.32	0.32
9465	0	0.32	1.00	0
9466	0.88	0.32	0	1.00

3.1.5 聚类分类

以上文梳理出的196个会话为对象进行聚类分类。以轮廓系数为评价标准,通过调整衰减系数 λ 获取最佳聚类结果。

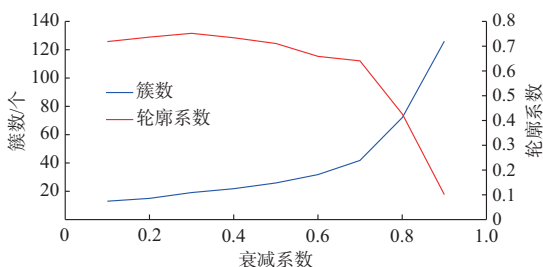


图5 最佳聚类结果分析
Fig. 5 Analysis of optimal clustering results

如图5所示,随着衰减系数 λ 的增大,聚类结果的簇数随之增加。轮廓系数在 $\lambda=0.3$ 时最大。随着聚类簇数的增大,单一点簇数增加,轮廓系数随之下降。综合来看, λ 选取0.3比较合适。

以衰减系数 $\lambda=0.3$ 时的聚类结果为例,获得聚类簇数19个,限于篇幅原因,这里简要展示聚类结果如表4所示,其中,会话集中的数字为该聚类包含的会话编号。

表4 聚类结果简要展示
Table 4 Brief display of clustering results

聚类	会话集
1	8655,8668,8856,8985,9061,9062,9067,9466
2	8884,8891,8961,9005,9022,9049,9109,9448,9466
⋮	⋮
18	9468,9450,8631,8632,8661,8699,9054,9073,9095,9131
19	9476,9478,8783,8799,8994,8996,9002,9004,9015

3.1.6 模式分析

依据前述操作模式挖掘结果,结合实际电网调

度业务工作情况,对相关模式进行分析。结果表明,挖掘的部分行为模式与调度员日常操作流程相吻合,验证了所提方法的有效性。

按照发生频率选取调度员经常进行的画面浏览行为模式挖掘结果,如表5所示。

表5 典型行为模式挖掘结果
Table 5 Mining results of typical behavior pattern

模式编号	行为模式
2	AGC-运行监控 .agc,电网负荷监视 .fac,联络线监视曲线 .sys,电网机组出力监视图 .sys
8	调度预案-首页 .sys,调度预案-断面越限 .sys,调度预案-独立网调 .sys,调度预案-独立计算本地 .sys
12	电网负荷监视 .fac,新能源受阻情况全景化展示 .sys,新能源监视图 .sys
14	电网负荷监视 .fac,龙羊峡水位曲线 .cur,水情信息监视 .sys

其中,行为模式2对应调度员日常工作中联络线守护的操作模式。在该业务场景中,调度员通过查看电网负荷监视、联络线监视曲线确定联络线是否运行在考核范围内,如果超出范围则需要查看AGC运行监控确认是否有机组可进行反向调节^[23]。

行为模式8为故障发生后进行故障预演的操作模式。调度员一般通过断面越限和独立计算等功能预演出故障发生后的断面状态数据,作为后续决策的参考依据。

行为模式12为调度员监视新能源出力受阻情况的模式,经常在上半发生。09:00—10:00处于新能源机组出力爬坡时段,此时调度员需要时常查看新能源受阻情况和新能源监视等图形监控新能源机组的运行情况,确保新能源的消纳能力^[24]。

行为模式14对应调度员进行调频、调峰的模式。水电站是比较经济高效的旋转备用电源,当调度员需要进行一次调频、深度调峰等操作时,通常需要查看直调水电站的水位曲线、出入库流量等数据,以判断水电站的调节能力边界^[25]。

3.2 推荐应用

基于本文的行为模式挖掘与操作推荐技术研发的调度操作智能推荐软件,为调度员的日常系统页面浏览和操作提供个性化、智能化推荐,整体推荐效果如附录A图A2所示。该系统以独立插件的方式与D5000系统人机界面实现集成,在使用时调度员点击调度系统页面的悬浮图标,推荐界面会以滑动窗口的方式在页面右侧显示,效果如图A2右侧推荐软件界面所示。

以行为模式2和14为例,说明智能操作推荐的效果。如附录A图A2所示,当调度员进入“电网负

荷监视图”后,智能操作推荐软件结合模式2自动推荐出调阅“AGC-运行监视”“联络线监视曲线”等关联界面,结合模式14自动推荐出龙羊峡等水电站水位曲线和“水情信息监视”等关联界面,点击按钮可以直接跳转到相应页面。

以此场景为例,简单对比使用推荐软件前后的操作效果差异。行为模式2原有系统从“电网负荷监视图”页面完成上述关联页面的跳转至少需要6步操作,分别为:1)打开“系统首页”;2)打开“实时监视”;3)打开“AGC-运行监视”;4)返回“系统首页”;5)打开“联络线考核主界面”;6)打开“联络线监视曲线”。而使用智能推荐功能只需要一键直接打开“AGC-运行监视”和“联络线监视曲线”画面,操作步骤由6步简化至2步,操作效率提升了约70%。

行为模式14原有系统至少需要经过5步操作才能完成上述操作,分别为:1)在龙羊峡水位数据上右键单击;2)在弹出的右键菜单上点击“今日曲线”选项,打开“龙羊峡水位曲线”画面;3)返回“系统首页”画面;4)打开“稳态监控”画面;5)点击水情信息按钮,打开“水情信息监视”画面。而使用智能推荐功能只需要一键直接打开“龙羊峡水位曲线”和“水情信息监视”画面,操作步骤由5步简化至2步,操作效率提升了约60%。

以上述调度会话为例,分析使用推荐工具带来的效率提升,如图6所示。未使用推荐工具时完成会话需要的操作步数为1778;使用推荐工具时完成会话需要的操作步数为1116。通过统计分析,使用推荐工具时可以减少约37.2%的页面跳转流程,显著提升了调度员在调度系统上的交互效率和便捷性。

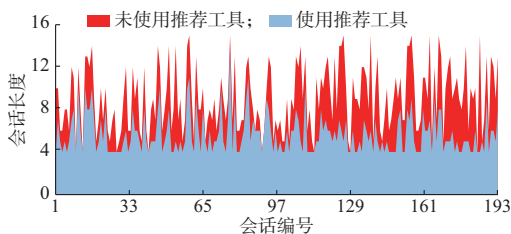


图6 工作效率对比
Fig. 6 Comparison of operation efficiency

此外,本软件系统还可提供依据操作热点统计的操作排行推荐、依据固定时段操作模式的定时操作推荐等多种操作推荐模式。例如对于行为模式12,推荐软件会在09:00—10:00直接定时推荐出当下调度员需要执行的打开“新能源监视图”、查看“新能源机组负荷曲线”等操作,无须频繁跳转到相关页面进行操作,通过对定时操作的智能提醒与一键交

互,可以高效地辅助调度员完成操作任务。

4 结语

本文主要阐述了基于调度员的历史操作数据,结合机器学习、数据分析等方法对调度操作行为模式进行挖掘与推荐的技术。通过在实际调度系统中的验证分析,表明挖掘出的调度员操作模式与实际电网调度业务具有高度一致性,验证了该技术的有效性。此外,该技术采用了FP-tree数据结构存储模式数据,并在此基础上通过在线匹配、置信评估等方法为调度员提供了准确的推荐服务和便捷的操作方式。

总的来说,调度员操作行为模式挖掘与推荐技术在改善调度员与调度系统交互方式、减轻调度员工作量、提高调度员工作效率等方面有显著的提升。但是本文提出的模式挖掘与操作推荐技术目前仍只适用于日常相对固化的流程性业务模式,在电网故障或异常等紧急场景下并不适用。当电网发生故障时,运行人员须结合电网运行数据、影响情况和处置经验进行针对性的处置,并且处置过程的部分操作不是通过系统而是通过电话或其他手段实施的,完整的处置记录很难准确获取。因此,在上述紧急场景下,如何结合电网运行情况和调度业务数据的复杂场景进行操作推荐,是后续研究的方向。

附录见本刊网络版(<http://www.aeps-info.com/aeps/ch/index.aspx>),扫英文摘要后二维码可以阅读网络全文。

参考文献

- [1] 胡彬,李苏苏.贵州电网调度优化操作模式建议[J].贵州电力技术,2012,15(2):73-75.
HU Bin, LI Susu. Suggestion on Guizhou power grid dispatching optimization mode [J]. Guizhou Electric Power Technology, 2012, 15(2): 73-75.
- [2] 王轶禹,王晶,冯长有,等.电网系统保护在线监视研究及典型应用[J].电力工程技术,2019,38(5):63-70.
WANG Yiyu, WANG Jing, FENG Changyou, et al. Online monitoring for large power system protection and its optimization strategy [J]. Electric Power Engineering Technology, 2019, 38(5): 63-70.
- [3] 崔晓丹,李碧君,李威,等.设备月度检修计划给定下的电网运行方式优化方法[J].电力系统保护与控制,2016,44(9):102-107.
CUI Xiaodan, LI Bijun, LI Wei, et al. Operation mode optimization method of power grid with equipment monthly maintenance planning [J]. Power System Protection and Control, 2016, 44(9): 102-107.
- [4] 杨涌新.检修方式下基于风险评估和控制的地区电网运行方式

- 调整[J].电力系统装备,2020,14:95-96.
- YANG Yongxin. Adjustment analysis of regional power grid operation mode based on risk assessment and control under maintenance[J]. Electric Power System Equipment, 2020, 14: 95-96.
- [5] 许洪强,姚建国,南贵林,等.未来电网调度控制系统应用功能的新特征[J].电力系统自动化,2018,42(1):1-7.
- XU Hongqiang, YAO Jianguo, NAN Guilin, et al. New features of application function for future dispatching and control systems [J]. Automation of Electric Power Systems, 2018, 42(1): 1-7.
- [6] 李慧芬.基于日志挖掘的移动应用用户行为分析系统的构建[D].北京:北京交通大学,2016.
- LI Huifen. Construction of mobile application user behavior analysis system based on log mining [D]. Beijing: Beijing Jiaotong University, 2016.
- [7] 王春玲,李川,李想.基于Apriori算法的高校Web日志挖掘系统构建[J].中国林业教育,2019,37(2):22-26.
- WANG Chunling, LI Chuan, LI Xiang. Construction of college web log mining system based on Apriori algorithm[J]. Forestry Education in China, 2019, 37(2): 22-26.
- [8] 邵超,宋淑米.基于信任关系下用户兴趣偏好的协同过滤推荐算法[J].计算机科学,2021,1:240-245.
- SHAO Chao, SONG Shumi. Collaborative filtering recommendation algorithm based on user preference under trust relationship[J]. Computer Science, 2021, 1: 240-245.
- [9] 李建祥.智能网页推荐系统的相关技术研究[D].成都:西南交通大学,2004.
- LI Jianxiang. Related technology research on intelligent web page recommendation system [D]. Chengdu: Southwest Jiaotong University, 2004.
- [10] 孙中悦,顾爽,范志静.互联网推荐系统在数字期刊中的应用[J].科技与出版,2021,4:57-62.
- SUN Zhongyue, GU Shuang, FAN Zhijing. Application of internet recommendation system in digital periodicals [J]. Science-Technology & Publication, 2021, 4: 57-62.
- [11] 高建军.地区电网调度运行的操作制度和注意事项[J].装备制造技术,2012(5):168-169.
- GAO Jianjun. Regional grid dispatching operation system and matters needing attention [J]. Equipment Manufacturing Technology, 2012(5): 168-169.
- [12] 陈郑平,米为民,林静怀,等.电网调控操作智能助手方案探讨[J].电力系统自动化,2019,43(22):173-178.
- CHEN Zhengping, MI Weimin, LIN Jinghuai, et al. Discussion on intelligence assistant scheme of dispatching and control operation in power grid [J]. Automation of Electric Power Systems, 2019, 43(22): 173-178.
- [13] 乔骥,王新迎,闵睿,等.面向电网调度故障处理的知识图谱框架与关键技术初探[J].中国电机工程学报,2020,40(18):5837-5848.
- QIAO Ji, WANG Xinying, MIN Rui, et al. Framework and key technologies of knowledge-graph-based fault handling system in power grid[J]. Proceedings of the CSEE, 2020, 40(18): 5837-5848.
- [14] 余建明,王小海,张越.面向智能调控领域的知识图谱构建与应用[J].电力系统保护与控制,2020,48(3):29-35.
- YU Jianming, WANG Xiaohai, ZHANG Yue. Construction and application of knowledge graph for intelligent dispatching and control[J]. Power System Protection and Control, 2020, 48(3): 29-35.
- [15] 宦蕾.基于Web日志挖掘的页面推荐的研究[D].南京:南京理工大学,2009.
- HUAN Lei. Research on web page recommendation based on log mining [D]. Nanjing: Nanjing University of Science and Technology, 2009.
- [16] 周旭.基于Web日志挖掘的个性化推荐方法研究[D].武汉:武汉理工大学,2014.
- ZHOU Xu. Research on personalized recommendation based on web log mining [D]. Wuhan: Wuhan University of Technology, 2014.
- [17] 严奉华,刘建平,杨凡丁.改进的Web访问日志会话识别算法[J].计算机工程与设计,2008,29(22):5685-5687.
- YAN Fenghua, LIU Jianping, YANG Fanding. Improved method for session analysis in web log mining [J]. Computer Engineering and Design, 2008, 29(22): 5685-5687.
- [18] 魏联滨,王伟臣,李慧,等.基于AP聚类和鲁棒优化的电网规划灵活性评估[J].电力系统及其自动化学报,2020,32(3):99-106.
- WEI Lianbin, WANG Weichen, LI Hui, et al. Evaluation on grid planning flexibility based on affinity propagation clustering and robust optimization [J]. Proceedings of the CSU-EPSA, 2020, 32(3): 99-106.
- [19] TAN P N, STEINBACH M, KUMAR V.数据挖掘导论[M].北京:人民邮电出版社,2011.
- TAN P N, STEINBACH M, KUMAR V. Introduction to data mining[M]. Beijing: Post & Telecom Press, 2011.
- [20] 朱志国.Web用户使用模式与兴趣挖掘方法研究[D].大连:大连理工大学,2010.
- ZHU Zhiguo. Research on methods of mining web user' usage patterns and interests [D]. Dalian: Dalian University of Technology, 2010.
- [21] 王治和,王淑艳,杜辉.基于密度敏感距离的改进模糊C均值聚类算法[J].计算机工程,2021,47(5):88-96.
- WANG Zhihe, WANG Shuyan, DU Hui. Improved fuzzy C-means clustering algorithm based on density-sensitive distance [J]. Computer Engineering, 2021, 47(5): 88-96.
- [22] 孙鸿艳,吉根林.一种新的基于FP_Growth的频繁项目集并行挖掘算法[J].南京师大学报(自然科学版),2016,39(4):19-24.
- SUN Hongyan, JI Genlin. New parallel algorithm for mining frequent item sets based on FP_Growth[J]. Journal of Nanjing Normal University (Natural Science Edition), 2016, 39(4): 19-24.
- [23] 高宗和,陈刚,杨军峰,等.特高压互联电网联络线功率控制[J].电力系统自动化,2009,33(15):51-55.
- GAO Zonghe, CHEN Gang, YANG Junfeng, et al. Active power control for tie lines in UHV inter connected power grid part on AGC control strategies [J]. Automation of Electric Power Systems, 2009, 33(15): 51-55.
- [24] 林俐,李北辰,孙勇,等.基于高比例新能源消纳的抽水蓄能容

量多时间尺度迭代优化配置模型[J]. 电网与清洁能源, 2021, 37(1): 104-111.

LIN Li, LI Beichen, SUN Yong, et al. Multi-time-scale iterative optimal configuration model of pumped storage capacity based on accommodation of high share new energy [J]. Power System and Clean Energy, 2021, 37(1): 104-111.

[25] 钟儒鸿, 程春田, 廖胜利. 兼顾多电网调峰与水电消纳的跨流域梯级水电站调度方法[J]. 电力系统自动化, 2021, 45(14): 114-122.

ZHONG Ruhong, CHENG Chuntian, LIAO Shengli. Inter-basin scheduling method of cascaded hydropower plants considering multi-grid peak shaving and hydropower accommodation [J]. Automation of Electric Power Systems,

2021, 45(14): 114-122.

吴自博(1991—), 男, 通信作者, 硕士, 主要研究方向: 大数据技术在电力调度系统中的研究与应用。E-mail: wuzibo2019@163.com

王 波(1981—), 男, 硕士, 高级工程师, 主要研究方向: 人工智能在调度自动化技术中的研究与应用开发。E-mail: wwwbo1981@163.com

陈 清(1994—), 男, 硕士, 主要研究方向: 人工智能在调度自动化技术中的研究与应用开发。E-mail: small_persimmon@163.com

(编辑 王梦岩)

Research and Application of Dispatch Operation Behavior Mining and Recommendation Technologies Based on Machine Learning

WU Zibo^{1,2}, WANG Bo^{1,2,3}, CHEN Qing^{1,2}, GUO Yaosong^{1,2}, ZHAO Jinghu^{1,2}, SHAN Xin^{1,2,3}

(1. NARI Group Corporation (State Grid Electric Power Research Institute), Nanjing 211106, China;

2. NARI Technology Co., Ltd., Nanjing 211106, China;

3. State Key Laboratory of Smart Grid Protection and Control, Nanjing 211106, China)

Abstract: Inspired by the Internet user recommendation technology, an effective dispatch operation behavior pattern mining and recommendation technology is proposed based on the historical operation records of operators in the dispatch system, combined with data analysis and machine learning algorithms. Firstly, the proposed technology collects and records the daily operation data of dispatchers and performs the data preprocessing through page type division, session recognition, and routine identification. Then, the behavior pattern mining of dispatchers is completed by using pattern analysis and cluster analysis, and the pattern data is compressed to a frequent pattern tree (FP-tree) by association rules. Finally, the real-time recommendation system based on the pattern tree structure is used to provide the operation recommendation services for dispatchers. The results of the verification and application in the actual dispatch system show that the proposed technology can effectively identify the operation behavior pattern of dispatchers and provide more accurate operation recommendation for dispatchers, thereby improving the interaction efficiency between the dispatcher and the dispatch system, and reducing the workload of dispatchers.

Key words: power grid dispatch; machine learning; behavior pattern; operation log; clustering mining; operation recommendation

