

doi: 10.3969/j.issn.0490-6756.2019.04.010

基于矩阵分解的电力物资需求预测

王竹君¹, 朱颖琪², 孙界平³

(1. 贵州电网有限责任公司贵阳供电局物流服务中心, 贵阳 550001;

2. 贵州电网有限责任公司信息中心, 贵阳 550000; 3. 四川大学计算机学院, 成都 610065)

摘要: 准确预测变电站及配网工程的物资需求, 对于节约工程成本, 提高资金利用率, 具有重要意义。尽管研究者在电力物资需求预测方面已经开展了一系列的研究, 提出了很多预测模型和算法, 例如基于神经网络的算法, 然而, 这些算法普遍存在输入数据过于简单和理想、仅对少数几种物资的需求量进行了预测实验、预测的准确率偏低等不足。因此, 目前生产系统普遍采用人工方式进行电力物资需求预测, 由有经验的领域专家根据工程初步设计方案预测各类物资的需求量。为了解决现有电力物资需求预测方法存在的不足, 本文提出基于矩阵分解的预测方法, 以电网建设项目物资需求历史数据和项目计划的部分物资作为输入, 通过矩阵分解算法对项目其他物资需求用量进行预测。矩阵分解算法不需要大量的历史用量数据, 只用部分项目的物资数据就能进行预测, 且算法不需要提前进行训练。

关键词: 电力物资; 需求预测; 矩阵分解

中图分类号: TP391 **文献标识码:** A **文章编号:** 0490-6756(2019)04-0639-06

Prediction of power grid material demand based on based on matrix decomposition

WANG Zhu-Jun¹, ZHU Ying-Qi², SUN Jie-Ping³

(1. Logistics Service Center of Guiyang Power Supply Bureau, Guizhou Power Grid Limited Liability Company, Guiyang 550001, China; 2. The Information Center of Guizhou Power Grid Limited Liability Company, Guiyang 550000, China; 3. College of Computer Science, Sichuan University, Chengdu 610065, China)

Abstract: It is of great significance to accurately predict the material demand of substation and distribution network for saving project cost and improving capital utilization. Researchers have carried out a series of studies on power material demand forecasting, and proposed many prediction models and algorithms such as neural network-based algorithms. However, these algorithms have several disadvantages. Specifically, these algorithms can only process simple and ideal input, predict the demand of limited materials, and suffer from the problem of low accuracy. As a result, in current production systems, the demand of electric power materials is predicted by experienced experts according to the preliminary design scheme of the project manually. In order to solve the existing shortcomings of the current demand forecasting methods, this paper proposes a forecasting method based on matrix decomposition. The method takes the historical data of the power grid construction project material requirements and part of the project plan as input, and use matrix decomposition algorithm to predict the demand for other materials in the project. The matrix decomposition algorithm can be implemented with the material data of

收稿日期: 2018-11-27

基金项目: 南方电网公司科技项目(GZKJXM20170162); 2018 四川省新一代人工智能重大专项(18ZDX0137)

作者简介: 王竹君(1988-), 女, 硕士, 经济师, 研究方向为电力物资管理及预测。

通讯作者: 孙界平. E-mail: sunjieping@scu.edu.cn

some projects instead of a large amount of historical usage data. In addition, the developed model does not need to be trained in advance.

Keywords: Power material; Demand forecasting; Matrix decomposition

1 引言

当前,我国社会经济快速发展,对电力能源的需求无论从量还是质上都提出了更高的要求.这些要求一方面促进了电网工程(变电站及配网工程)市场的繁荣,另一方面也对相关企业提出了更大的挑战.相关企业只有借助于高科技尤其是人工智能技术,优化企业管理和各种资源配置,提高资源利用率和工程设计开发效率,才能适应市场新情况,从容应对这些新的更大的挑战.在众多挑战中,如何准确预测变电站及配网工程的物资需求,从而在保障工程进度的前提下提高资金的利用率,节约成本,对于电力企业新的市场环境下快速发展,具有重要意义.

在电力物资需求预测方面,研究者已经开展了一系列的研究工作^[1,2],提出了很多预测模型和算法^[3-8],例如文献[3]提出了基于神经网络的预测算法.然而,这些算法以历史数据中有限的几维作为输入(自变量),与实际应用需求相距甚远.在实际应用中,一个电力工程需要的物资种类有几万种,而现有的研究仅在少数几种物资上进行了预测实验.所以现有的研究成果已远远不能满足实际需求.因此,目前的生产系统仍然普遍采用人工方式进行电力物资需求预测,由有经验的领域专家根据工程初步设计方案预测各类物资的需求量.

为了解决现有电力物资需求预测存在的问题,本文提出了基于矩阵分解的预测算法,利用电网建设项目物资需求历史数据设计进行实验,取得了较好的效果.本文后续内容组织如下:第 2 节介绍相关工作,分析现有电力物资需求预测研究成果和不足;第 3 节提出基于矩阵分解的电力物资需求预测算法,给出算法框架;第 4 节详细介绍算法的细节,包括数据预处理和基于矩阵分解算法预测物资用量;第 5 节介绍了实验设置及实验结果;最后第 6 节对全文工作进行总结,讨论本文工作的后续改进和发展方向.

2 相关研究

电力物资需求预测具有重要的应用价值,因此引起了学术界的广泛重视.截至目前,学者们已经

提出了一些模型和算法^[3-8]用于解决电力物资需求预测问题.文献[1]在分析电网物资需求特性的基础上,提出了利用改进 BP 神经网络进行预测的方法,该方法用 SCG(Scaled Conjugate Gradient)和遗传算法分别对标准 BP 训练算法本身和权值初始化、阈值选择方法进行了优化,通过 110 kV 新建线路工程的钢芯铝绞线需求预测对改进方法进行了实验验证.文献[3]设计的 BP 网络只包含一个隐藏层,输入向量的维数为 2,分别表示线路长度和地形因素,输出层神经元数是 1,其输出值为预测的钢芯铝绞线数量.作者还通过实验检验了不同的隐藏层单元数量对预测准确率的影响.文献[4]采用支持向量机 SVM(Support Vector Machine)回归来解决电网物资需求预测问题,已增加了混沌搜索算子的人工鱼群算法对支持向量机参数和核函数(即支持向量机的超参数)进行优化搜索.文献[4]未说明输入向量的维数,并且仅给出了控制电缆和钢芯铝绞线两种物资的预测结果.文献[5]研究了物资需求自动审核的问题,考虑到电力行业上报的物资需求记录具有不完全结构化、短文本的特点,物资需求审核被归结为文本分类问题,采用的分类模型是 SVM.文献[6]提出了基于多元回归分析法的电网应急物资需求预测方法,但考虑的自变量仅有三个,分别为地区风力等级、降雨量和年度用电总量,并且未给出实验结果,也未说明对什么类型的物资进行需求预测,还是对所有的物资都是利用这三个自变量进行需求预测.文献[7]过计算实例简单讨论了指数平滑等方法在电力物资需求预测方面的可能应用.文献[8]利用关联规则算法研究物料间的频繁项集,从而得到物料之间的相关关系.

综上所述,尽管在电力物资需求预测方面已经进行了一些研究和尝试,但是普遍存在的问题是实用性较差,预测所依据的数据过于理想化,均为少数几个属性表达的结构化数据;其次,能够有效预测的物资种类有限.电网工程建设所需要的物资种类多达几万种,它们都属于被预测的对象,仅对几种物资进行预测不具有实用性.

另外,电网工程物资种类繁多,而单个电网项目所用物资相比很少,利用历史物资用量进行预测

时,存在数据稀疏问题。

针对现有研究成果存在的不足以及电网项目物资需求的特点,本文将研究基于矩阵分解^[9]的电网工程物资需求预测,提出基于矩阵分解的预测算法。矩阵分解,是推荐系统中应用最为广泛的技术之一^[10-13]。推荐系统分析用户兴趣,在用户群中找到指定用户的相似(兴趣)用户,综合这些相似用户对某一信息的评价,形成系统对该指定用户对此信息的喜好程度预测。在电网工程项目物资需求预测

中,可把每个项目看作用户,某种物资用量看作用户对某信息的喜好程度,利用矩阵分解的思想,可对项目的未知物资用量进行预测。

3 预测算法

本文提出的算法以电网建设项目物资需求历史数据和项目部分物资计划用量为输入,项目其他物资需求量的预测值为输出。算法框架如图 1 所示。

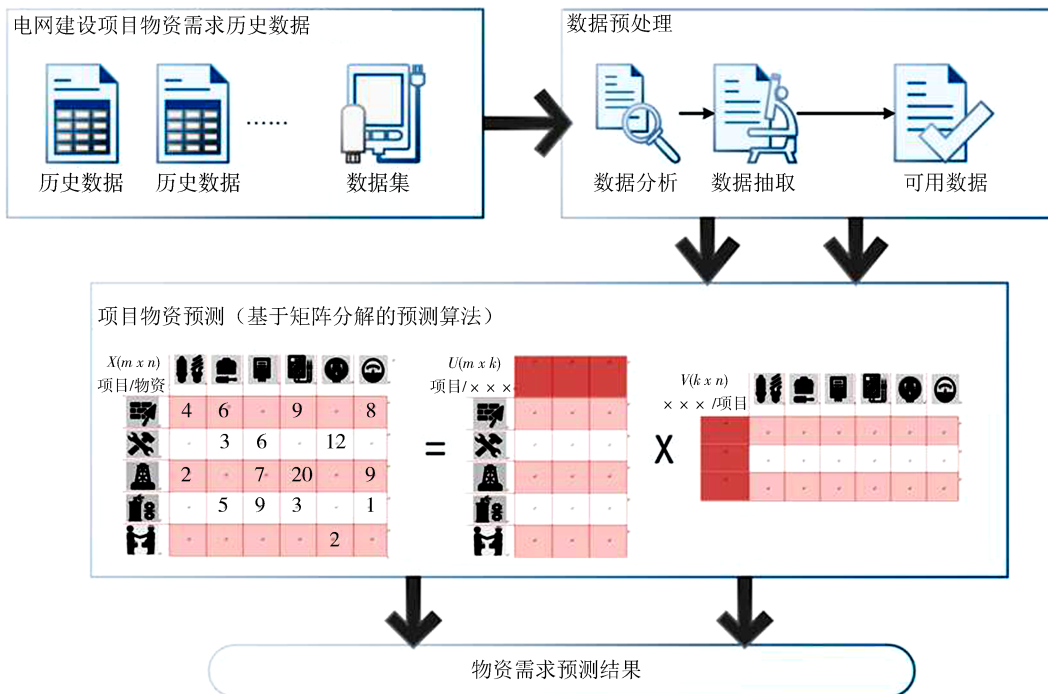


图 1 算法框架

Fig. 1 Algorithm framework

图 1 中,算法主要分为两个步骤:数据预处理和项目物资预测。我们所获取到的电网工程项目物资需求历史用量数据需要经过数据预处理后,模型才能利用这些历史数据进行预测。由于电网工程项目物资种类繁多,有几万种,而对于一个项目,每一种类的物资都可能用到,直接预测几万种物资用量是不现实的。因此,设计了数据预处理过程,对几万种物资进行一定泛化处理,得到所要预测的物资种类。在数据预处理后,利用基于矩阵分解算法^[9,14,15]进行项目物资预测。该算法将历史项目数量 M 和物资种类 N 看作矩阵 $A(M \times N)$,则矩阵的每一个元素 a_{ij} 表示第 i 个项目的第 j 类物资的用量。通过将矩阵分解 A 分解为两个新的矩阵 P 和 Q ,对于矩阵 A 中未知的数据,则可通过 P 和 Q 得到。具体算法将在下一节中详细介绍。

4 基于矩阵分解的电网物资预测

4.1 预处理过程

数据是从贵阳电网的系统导出的电网工程项目物资需求历史用量数据。系统导出的数据以表格的形式存储,这其中包含了许多对预测无用的信息,如交货日期、采购级别、推荐供货商、采购方式和采购完成时间等。在数据预处理阶段首要任务是分析处理获取到对预测有帮助的数据如项目名称、物资用量、物资型号和物资用量单位等数据。

其次,在电网工程项目中的物资种类高达几万种,而在单个的电网项目中,每一种物资都有用到的可能,直接对这几万种物资预测其需求用量是很困难的。且在单个的电网项目中用到的物资是有限的,相比于几万种物资类别是非常少的,在利用历

史物资需求用量数据时,存在着数据稀疏问题,所以我们采用向上泛化的思想,即我们查找每个物资的父节点,然后再判断每个父节点下面对应的物资的单位是否一致,如果一致,合并所有子节点统一用父节点物资名称,并继续向上查找,如果不一致则就停止.并且如果一个项目中出现多个父节点名称一样,我们就进行叠加运算,这样在一定程度上缓解了数据稀疏问题.进行泛化之后,物资种类得到了一定的减少.

4.2 基于矩阵分解的数据补全算法

在推荐系统中,常常遇到这样的问题,我们有很多用户和物品,也有少部分用户对少部分物品的评分,我们希望预测目标用户对其他未评分物品的评分,进而将评分高的物品推荐给目标用户.正如电网项目物资需求预测问题中,我们有许多项目和物资类别,也有部分项目中部分物资的用量,我们希望预测目标项目对其他物资的需求用量.如表 1 的项目物资表,算法的任务就是预测出项目 4 和项目 5 中没有值的物资用量.如果将 m 个项目和 n 种物资对应的用量看做一个矩阵 M ,可用矩阵分解解决该问题.矩阵分解是推荐系统中广泛使用的方法.基于矩阵分解的数据补全算法的伪代码如算法 1 所示.

算法 1 基于矩阵分解的数据补全算法

输入: 参数 M ; 需要进行分解的矩阵 $M_{m \times n}$;

参数 k : 分解的隐变量的个数

输出: 分解得到的矩阵 $P_{m \times k}$ 和 $Q_{k \times n}$

Initialize P , Q and α , λ , error;

α is leaning rate, while λ is regularization coefficient

for t in range (10000): # 设置迭代次数

for i in range(m):

for j in range(n):

if math.fabs($M[i][j]$)! = 0. :

err= $M[i][j]$ - dot($P[i, :]$, $Q[:, j]$) # 计算

误差

error=math.pow(err, 2) # 计算误差平方

for r in range(0, k):

$P[i][r]$ = $P[i][r]$ + $\alpha * (\text{err} * Q[j][r]$ -

$\lambda * P[i][r])$ # 更新 P

$Q[j][r]$ = $Q[j][r]$ + $\alpha * (\text{err} * P[i][r]$ -

$\lambda * Q[j][r])$ # 更新 Q

return [P , Q]

表 1 项目物资表

Tab. 1 Project material list

项目/物资	物资 1	物资 2	物资 3	物资 4	物资 5	物资 6
项目 1	4	6	0	9	0	8
项目 2	0	3	6	0	12	0
项目 3	2	0	7	20	0	9
项目 4		5	9			1
项目 5					2	

矩阵分解^[9,14]指将一个矩阵分解成两个或者多个矩阵的乘积.对于表 1 的项目物资所构成的矩阵项目物资用量矩阵,记为 $M_{m \times n}$.可以将其分解成两个或者多个矩阵的乘积,本算法中分解成两个矩阵 $P_{m \times k}$ 和 $Q_{k \times n}$,使得矩阵 $P_{m \times k}$ 和 $Q_{k \times n}$ 的乘积能够还原原始的矩阵 $M_{m \times n}$,如式(1).

$$M_{m \times n} \approx P_{m \times k} \times Q_{k \times n} = \hat{M}_{m \times n} \quad (1)$$

如何得到矩阵 $P_{m \times k}$ 和 $Q_{k \times n}$,可以将这个问题转化成机器学习中的回归问题进行求解.要使原始的物资用量矩阵 $M_{m \times n}$ 与新构建的矩阵 $\hat{M}_{m \times n}$ 尽可能接近,误差尽可能小,采用均方误差作为损失函数,即式(2)所示.

$$e_{i,j}^2 = (m_{i,j} - \hat{m}_{i,j})^2 = (m_{i,j} - \sum_{k=1}^k p_{i,k} q_{k,j})^2 \quad (2)$$

最后需要求解的目标函数 L ,如下式.

$$L = \min \text{loss} = \sum_{i,j} e_{i,j}^2 \quad (3)$$

在实际应用中,为了防止过拟合,会加入一个 L_2 的正则化项^[16],如下式.

$$L' = \sum_{i,j} (m_{i,j} - \sum_{k=1}^k p_{i,k} q_{k,j})^2 + \lambda \sum_{k=1}^k (p_{i,k}^2 + q_{k,j}^2) \quad (4)$$

其中, λ 为正则化系数.对于这个优化问题,通过随机梯度下降法来进行求解.

将式(4)分别对 $p_{i,k}$, $q_{j,k}$ 求导我们得到式(5)和式(6).

$$\frac{\partial L}{\partial p_{i,k}} = -2(m_{i,j} - \sum_{k=1}^k p_{i,k} q_{k,j}) q_{k,j} + 2\lambda p_{i,k} \quad (5)$$

$$\frac{\partial L}{\partial q_{k,j}} = -2(m_{i,j} - \sum_{k=1}^k p_{i,k} q_{k,j}) p_{i,k} + 2\lambda q_{k,j} \quad (6)$$

由式(5)和式(6)得,在随机梯度下降法迭代时 $p_{i,k}$, $q_{j,k}$ 的迭代公式为式(7)和式(8).

$$p_{i,k} = p_{i,k} + \alpha((m_{i,j} - \sum_{k=1}^k p_{i,k}q_{k,j})q_{k,j} + \lambda p_{i,k}) \quad (7)$$

$$q_{k,j} = q_{k,j} + \alpha((m_{i,j} - \sum_{k=1}^k p_{i,k}q_{k,j})p_{i,k} + \lambda q_{k,j}) \quad (8)$$

通过迭代我们最终可以将矩阵 $M_{m \times n}$ 分解得到 $P_{m \times k}$ 和 $Q_{k \times n}$, 对于项目 i 的第 j 个物资的物资预测用量 $\hat{m}_{i,j,k}$ 可通过式(9)计算.

$$\hat{m}_{i,j} = \sum_{k=1}^k p_{i,k}q_{k,j} \quad (9)$$

5 实验结果及分析

5.1 实验设置

实验数据是从贵阳电网的系统导出的电网工程项目物资需求历史用量数据. 结合项目需求, 我们首先筛选出项目编号中包含 WN 和 WP 的项目. 由于物资类别过多, 以及电网项目物资历史用量数据稀疏, 所以我们采用向上泛化的思想, 合并了部分物资. 然后我们统计出物资在所有项目中出现的次数, 进行排序, 分别选取前 10, 20, 30, 50, 70 的物资, 筛选出对应的项目进行预测.

在实验时, 针对前 10, 20, 30, 50, 70 的物资分别选取了 60 个项目, 每次实验时把其中一个项目的某一物资的值去掉作为预测目标. 因为不同项目有着不同的数量级, 所以预测结果采用的评价标准是相对误差. 每一个类别的物资都选取了十个项目进行该物资的用量预测, 最后得到平均相对误差.

实验中, 设置梯度下降的迭代过程的学习率 $\alpha=0.00001$, 正则项系数 $\lambda=0.001$. 为分解不同维度的项目物资用量矩阵设置不同的 K 值, 这里的 K 指的是项目物资用量矩阵 $M_{m \times n}$ 分解成的两个矩阵 $P_{m \times k}$ 和 $Q_{k \times n}$ 的行和列数. 本实验的 K 值设置如表 2.

表 5 部分物资预测结果

Tab. 5 Part of material forecast results

用量	物资 1	物资 2	物资 3	物资 4	物资 5	物资 6	物资 7	物资 8	物资 9	物资 10
实际用量	23.295	3.493	250	366	190	232	166	112	166	268
预测用量	16.74	3.073	49.5	212.4	252.4	220	222.4	346	191	272
相对误差(%)	28.1	12.0	81.1	41.9	32.8	5.1	33.9	209	15	1.5

6 结 论

针对电力物资需求的特殊性 & 现有研究成果存在的不足, 本文提出了基于矩阵分解算法预测电

表 2 参数设置

Tab. 2 Parameter settings

物资类别数	10	20	30	50	70
K	3	5	10	15	20

5.2 实验结果

实验结果如表 3 所示. 从实验结果可以发现随着物资类别数目增加, 平均相对误差在逐渐增大. 导致这一原因的可能是随着物资类别的增加, 项目物资用量矩阵中的零也会随之增加, 零元素过多, 矩阵分解效果不理想. 其次在项目物资用量矩阵(如表 4)中, 同一物资在不同项目中的用量差别大, 在同一项目中不同物资用量也相差甚远, 这也是导致实验结果不够理想的原因.

表 3 实验结果

Tab. 3 Experimental results

物资类别数	10	20	30	50	70
平均相对误差(%)	31.5	44.3	67.3	74.8	86.5

表 4 部分项目物资用量矩阵

Tab. 4 Part of the project material matrix

项目	物资 1	物资 2	物资 3	物资 4	物资 5
项目 1	693.4836	11.869	1787	11597	12882
项目 2	153.9202	0.48	761	2647	3344
项目 3	62.578	0.75	313	2048	2518

表 5 是对 60×10 的物资用量矩阵分解后所得到的预测结果. 可以看到对于物资 6 和物资 10, 本文所提出的算法能比较准确的预测, 相对误差在 5% 以内. 但物资 8 和物资 3 预测出的结果与实际用量相差很大. 造成预测结果不理想的原因是这两类物资需要更多的历史用量数据发现其规律, 本实验矩阵分解所利用的项目过少, 仅用了 60 个.

网项目物资需求用量. 相对于已有的研究成果, 本文提出的算法不需要大量的历史用量数据, 只用部分项目的物资数据就能进行预测; 也不需要训练, 算法简单. 算法能够利用电网项目物资需求历

史用量和部分项目物资计划用量,预测出其他物资的用量,这有利于为物资需求用量计划提供参考,同时对节约物资成本具有重要意义。

参考文献:

- [1] 毕子健,王翎颖. 电网物资需求预测方法研究 [J]. 华北电力技术, 2015(10): 26.
- [2] 陈啸,熊国宏,徐世阳. 基于大数据的电力采购物资质量管理体系的研究与探索(英文) [J]. 机床与液压, 2017(06): 132.
- [3] 沈男. 基于改进 BP 神经网络的电网物资需求预测研究 [D]. 北京: 华北电力大学, 2014.
- [4] 宋斌. 应用人工鱼群算法研究电网物资需求预测问题 [J]. 管理纵横, 2013(29): 178.
- [5] 韩戟,何成浩,苏星,等. 一种基于 SVM 的电力行业物资需求预测方法 [J]. 电气技术, 2016(12): 152.
- [6] 睦楷,王语涵,王少勇,等. 基于多元回归分析法的电网应急物资需求预测方法 [J]. 电子技术与软件工程, 2016(23): 195.
- [7] 谢荣伟. 基于需求特性分类的电力物资需求预测方法研究 [J]. 价值工程, 2016, 35: 57.
- [8] 温富国,许斌,李金霞. 基于关联规则的电网物资相关性研究 [J]. 电子世界, 2018, 543: 109.
- [9] 李乐,章毓晋. 非负矩阵分解算法综述 [J]. 电子学报, 2008, 36: 737.
- [10] 黄贤英,龙姝言,谢晋. 基于用户非对称相似性的协同过滤推荐算法 [J]. 四川大学学报: 自然科学版, 2018, 55: 489.
- [11] 金玉,崔兰兰,孙界平,等. 基于综合相似度迁移的协同过滤算法 [J]. 四川大学学报: 自然科学版, 2018, 55: 477.
- [12] 刘慧婷,陈艳,肖慧慧. 基于用户偏好的矩阵分解推荐算法 [J]. 计算机应用, 2015(s2): 118.
- [13] 杨阳,向阳,熊磊. 基于矩阵分解与用户近邻模型的协同过滤推荐算法 [J]. 计算机应用, 2012, 32: 395.
- [14] Gai L, Lei L, Polytechnic S. Collaborative filtering algorithm based on matrix decomposition [J]. Comput Eng Appl, 2011, 47: 1.
- [15] Li J. Research on collaborative filtering recommendation algorithm based on matrix decomposition method [C]// International Conference on Mechanics, Materials, Chemistry and Computer Engineering. [s. l.]:[s. n.], 2015.
- [16] 杜世强,石玉清,马明,王维兰. L 3/2 正则化图非负矩阵分解算法 [J]. 吉林大学学报: 理学版, 2014, 52: 1007.

引用本文格式:

- 中文: 王竹君,朱颖琪,孙界平. 基于矩阵分解的电力物资需求预测 [J]. 四川大学学报: 自然科学版, 2019, 56: 639.
- 英文: Wang Z J, Zhu Y Q, Sun J P. Prediction of power grid material demand based on based on matrix decomposition [J]. J Sichuan Univ: Nat Sci Ed, 2019, 56: 639.