

Research on Forecasting Stock Demand Based on Big Data Analysis

Yonghuan Hu¹, Fanghua Hong², Feng Xiao³, Fengna Dong³, Yongxu Zhang³

¹State Grid Shanghai Company, Shanghai

²State Grid Shanghai Company Material Company, Shanghai

³Shanghai Jiulong Enterprise Management Consulting Co. Ltd., Shanghai

Email: fengnadong@126.com

Received: Feb. 2nd, 2019; accepted: Feb. 17th, 2019; published: Feb. 25th, 2019

Abstract

Under the background of material intensification and intelligent supply chain management, this paper builds a set of top-down and bottom-up prediction models based on protocol inventory demand analysis, application system engineering theory and modern demand chain management theory. Based on the predictive model, this paper establishes a procurement demand forecasting system suitable for the procurement demand characteristics of the power system, and applies a large amount of historical data to analyze the forecasting effect for different life cycle stages and different demand types. In the analysis process, this paper improves the prediction accuracy by continuously adjusting the model parameters, and in the process of continuously improving the demand forecasting system, solidifies the top-down forecasting model, and finally builds a large planning system based on the project's full-process visualization platform.

Keywords

Big Data Analysis, Agreement Stock, Demand Forecast

基于大数据分析的协议库存需求预测研究

胡永焕¹, 洪芳华², 肖 锋³, 董凤娜³, 张永旭³

¹国网上海市电力公司, 上海

²国网上海市电力公司物资公司, 上海

³上海久隆企业管理咨询有限公司, 上海

Email: fengnadong@126.com

摘要

在物资集约化和供应链管理智能化的背景下，本文基于协议库存需求分析应用系统工程理论和现代需求链管理理论，构建了包括自上而下和自下而上两种预测模型，并根据预测模型建立适合电力系统采购需求特性的采购需求预测体系，并应用大量历史数据针对不同生命周期阶段和不同的需求类型来分析预测效果，分析过程中通过不断调整模型参数以提高预测精准度，在不断完善需求预测体系的过程中，固化自上而下预测模型，最终基于项目全程可视化平台，打造大计划体系。

关键词

大数据分析，协议库存，需求预测

Copyright © 2019 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

1.1. 研究背景与意义

协议库存即由需求方根据物资需求单位未来一定时期内物资需求的预测，应用统一的物料编码及技术规范，将所需的物资分类汇总，通过招标或其他采购方式确定供应商，并与其签订框架采购协议，待具体物资需求申请产生后，直接按照协议分配供应商，相当于采购货物已经提前在仓库备好，可以缩短采购供应周期，为工程建设争取时间，同时可减少需求方自身的库存，节约运营成本[1]。

随着物资集约化管理工作不断深入，协议库存采购在电力物资采购中所占的比例逐年增加，采购范围不断扩大，对协议库存物资的需求计划管理提出挑战，同时对协议库存需求预测准确度的要求越来越高。

本文通过大数据采购需求预测模块的建设，开展采购链全流程建模，推动电力物资全供应链的统一数据模型的构建，开发一套适用于电力物资采购历史状况和未来发展的数据预处理算法和面向电力物资的大数据定制化预测封装算法，并通过完成大数据需求预测模块的实施，包含数据预处理、采购需求分析、采购需求预测，持续推动物资环节的深化应用，提升用于采购目录和需求计划编制的的数据质量，探索物资需求计划的专业数据变化趋势、特征及规律，为物资管理效率效益提升提供决策支撑[2]。

1.2. 研究思路与目标

在基于需求分析的基础上，应用系统工程理论和现代需求链管理理论，构建了一套科学完整的适合电力公司需求特性的采购需求预测体系，主要分为自下而上需求预测和自上而下需求预测两种预测模型，如图1所示。

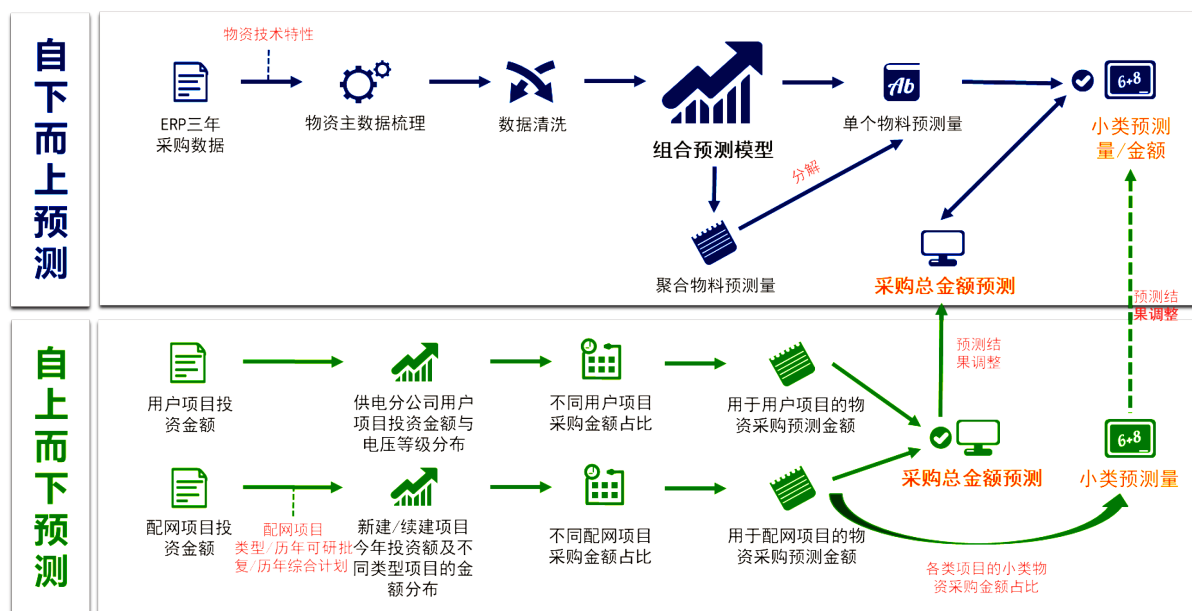


Figure 1. Big data procurement demand forecasting model
图 1. 大数据采购需求预测模型思路图

基于研究形成的大数据需求预测应用，在管理机制上，物资计划管理要与前端项目需求单位建立互联融合机制、与后端供应建立协同管控机制，实现物资需求计划的自动预警、辅助提报和对供应保障的统筹协同，提升需求计划准确性。

2. 协议库存需求预测体系

协议库存需求预测体系由数据清洗、采购物资需求特性分析和采购物资需求预测模型设计组成。

2.1. 数据清洗

为了保证数据分析过程中数据的可靠性，需对原始数据进行预处理。原始信息的搜集、整理、清洗、分析的过程，是进行需求预测的前提和基础。数据预处理全过程中最主要包括冲销处理、单价异常处理、去除季节性异常处理和负值处理四个过程。

1) 冲销处理

采购数据中有很多属于冲销数据即历史物资消耗记录存在反复冲销问题。如果不进行处理会对实际采购数据有很大的影响，因此在数据分析之前，需要把冲销数据去除。

步骤一：寻找所有采购数据为正的采购信息，并提取出匹配信息：数量、本位币金额、物料编码、过账日期、库位；

步骤二：在匹配时间范围内(向前 12 个月，向后 1 个月)查找满足匹配条件(物料编码相同、数量相反、金额在正负 10%之内、库位相同)的数据。

步骤三：将已匹配的数据数量修改成 0，本位币金额不做修改。

2) 单价异常处理

企业会出于一些实际情况的考虑，在保证金额准确的前提下，对采购数量做一些修改，因此此条采购记录的单价就会偏离正常值，通过判断每条采购记录的单价，以此来判断这条采购信息是否合理，并在此基础上修正采购数量。

步骤一：通过对 2015 年到 2017 年的采购订单数据整理获得每个物资每次采购的单价，并四舍五入取整。

步骤二：统计每个单价的出现频率，并依据频率大小，从大到小排序。

步骤三：挑选出排名前五的单价和频率，对于某些物资的不同的单价个数少于 5 个的，只取这几个单价和频率即可。

步骤四：确定每个物资的中心单价、最高单价、最低单价。如果该物资频率最高的单价与同属一个小类的其他物资的中心单价在同一数量级上或者差别小于 50%，那么就将该单价作为中心单价，否则根据物料描述选择规格最近的物资平均中心单价作为该物质的中心单价；如果该物资挑选出来的单价最大值与同属一个小类的其他物资的最大单价在同一数量级上或者差别小于 50%，那么就将该单价作为最大单价，否则根据物料描述选择规格最近的物资平均最大单价作为该物质的最大单价；最小单价同最大单价计算方法。

步骤五：选择异常单价范围。根据经验和上海电力实际情况，我们将异常单价范围设置为小于最小单价的 10%和大于最大单价的 900%，即我们认为大于 10 倍最大单价和小于 1/10 倍最小单价是异常单价，是需要进行修正的。

步骤六：通过计算机选出符合条件的采购信息，并对采购数量进行修正。

3) 去除季节性异常处理

通过对部分数据观察、分析发现，部分物资采购具有明显的季节性。因此通过去除季节性来观测哪些月份出现了异动采购——即不常用物资偶尔发生的大规模非常规采购，并对异常值进行修正，这些异常的采购值可能是由于项目因素或者其他在这一阶段没有考虑的因素，并在分析中进一步挖掘。

步骤一：通过程序计算出每个物资每月的季节因子。

步骤二：通过将每月采购数量/对应月份的季节因子得到去除季节因子的历史值。

步骤三：选择异常值范围，根据上海电力的调研分析和行业经验，我们将大于去除季节因子的历史值的均值 + 3 倍标准差的采购量或者金额作为异常值。

步骤四：通过计算机查找在异常范围内采购信息，并做修正。修正的原则是将异常月份采购调整到合理的范围内即可。

4) 负值处理

对于某些物资会出现某个月份的采购量小于零(这里与电力系统里面采购为负，入库为正的情况正好相反)的这种情况，与实际需求的非负性矛盾，这是由于采购后退回引起，实际上没有产生需求，为剔除这种情况的影响，需将负值剔除。

步骤一：挑选出采购数量 X 或者金额为负采购月份。

步骤二：将负值 X 加到前一个月份的采购量 Y 上，该月采购为 0，前一个月份出口为 Y + X。

步骤三：如果 Y + X 大于零，则结束计算，如果 Y + X 小于零则再向前一个月去除负值，同步骤二。

2.2. 采购物资需求特性分析

一般来说，需求特征可以分为四种，即水平需求、间歇需求、季节性需求和周期需求，与之相关的需求分析主要包括：1) ABC 分类；2) 连续性分析(判断需求是否连续)；3) 波动性分析(判断需求的预测难度)；4) 季节性分析(判断需求是否具备显著季节性特征)；5) 生命周期分析(判断对象是否具备显著生命周期特点)。对于电力物资，由于所有的协议库存物资都处于生命周期的成熟阶段，因此主要考虑前四种分析。

① ABC 分类

ABC 分类法是库存管理中非常常用的方法,该方法通过对物资金额进行帕累托排列,将数量占比少而金额占比高的物资作为 A 类物资,提倡管理者重点关注这些物资的需求,获得较高的管理效用。其中,A 类产品指的是物料条目少但采购金额较高的物资,B 类产品指的是品种适中价值也适中的物资,C 类产品指的是品种多而价值低的物资。

② 连续性分析

电力物资在需求连续性方面需求呈现不同的特点,由于连续和间断的需求适用的预测模型不同,因此需要对每一个物资分析其连续性。

连续性分析主要通过分析物资的采购次数(即需求次数)来实现,通过统计物资每个月的采购情况,计算月平均需求间隔和月采购比例,并与设定的阈值比较判定物资的连续性。

③ 波动性分析

由于季节性需求、项目式需求或者其他一些因素导致需求围绕均线上下波动,这种波动性(平稳性)主要是通过物资的月采购量标准差与月采购均值之比(变异系数 CV)来衡量。

④ 季节性分析

季节性分析主要用于判断物资采购是否具有季节性规律。季节性主要根据物资月采购的季节性指数变化进行判定,季节指数可以简单的理解成各月度平均需求与全年平均需求的比值。

⑤ 相关性检验

通过将上述特性指标整理后对数据进行标准化并使用 SPSS 24.0 进行相关性分析,并从结果中选择有较强特征相关性的维度,基于物资需求特性进行标签划分,进而对每个物资从物料及小类维度进行组合预测,通过对不同模型预测结果进行分析,赋予不同的权重,能够有效提升需求计划的适应度。

2.3. 采购物资需求预测模型

2.3.1. 自下而上的采购需求预测模型

对于电力物资数据可采用同时处理连续性需求和间断性需求的多个预测模型进行组合预测,以实现预测模型选择的自动化和智能化。进入组合预测模型的单体预测模型包括,适合预测需求连续性强、季节波动性明显的 Winter 模型、适合需求趋势性明显的曲线拟合模型和适合需求间断的 Croston 模型,设计算法对于不同预测模型的预测结果进行主观赋权,赋予不同的权重,有效提升需求计划的适应度。具体预测模型如下:

改进 winter 模型: ① 计算历史采购的季节因子(月度指数) ② 剔除历史数据中季节因子,形成相对平滑的采购趋势; ③ 使用 6 种拟合曲线对采购趋势进行拟合,选择拟合优度最好的模型预测未来的趋势值; ④ 将趋势值乘上相应月份的季节因子得到最终的预测结果。

Winters 乘法模型: ① 计算初始季节因子,剔除季节因子的历史值以及初始趋势值; ② 选择最优的平滑系数分别对季节因子,剔除季节因子的历史值,趋势值进行指数平滑,得到最终的季节因子,剔除季节因子的历史值和趋势值; ③ 将剔除季节因子的历史值加上趋势值再乘上季节因子得到最终的预测值。

Croston 模型: ① 将原始时间序列分成非零需求序列和需求间隔序列; ② 选择最优平滑系数分别对非零需求和需求间隔进行指数平滑; ③ 得到最后一期的非零需求和需求间隔的指数平滑值,将预测的非零需求量平均到这段需求间隔内; ④ 更新非零需求和需求间隔,继续预测。

如图 2 所示,将三类预测模型整合到组合预测模型中对每一颗物料进行预测,该模型既可以自动为每颗物料选择最优预测模型,也可以进行多个模型的智能组合预测,这里的组合预测是指以一定的规则“每一个模型得到的需求预测值”赋予权重,使用加权平均值作为最终输出的组合预测结果,权重设定则按照“预测误差最小”的方式进行寻优计算。同时,允许多颗技术特性相近的物料进行合并预测,之

后再分解得到各颗物料的预测结果，该过程称之为聚合-分解预测过程，最终得到单一物料的“自下而上需求预测”结果。

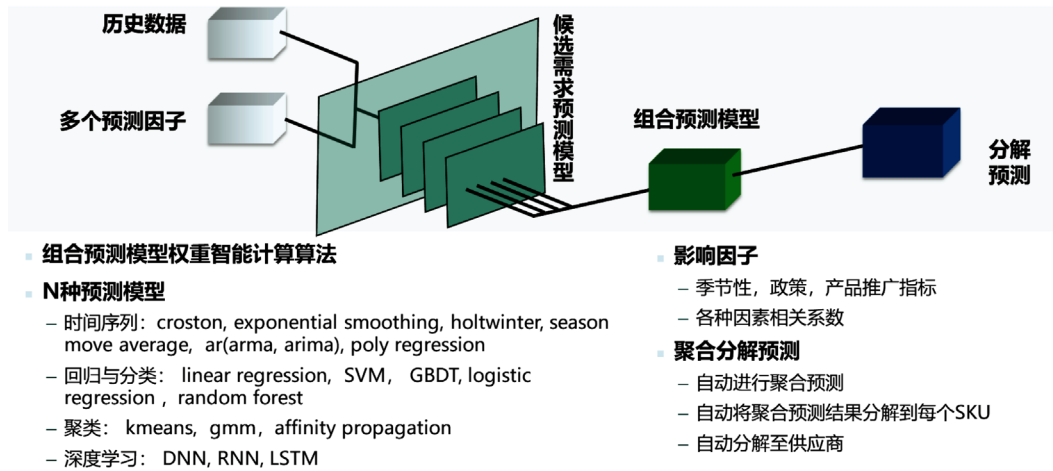


Figure 2. Bottom-up demand forecasting model framework for power materials
图 2. 电力物资自下而上需求预测模型框架

2.3.2. 自上而下的采购需求预测模型

如图 3 所示的电力物资自上而下的采购需求预测模型，主要流程为：

- 1) 通过对近三年可研批复数据和综合计划数据以及项目物资历史采购订单的整理，得到项目实际计划采购金额、实际采购金额、电压等级以及相关小类采购情况等属性信息；
- 2) 通过项目的属性信息对项目进行聚类，并确认聚类结果所属的项目类型；
- 3) 以总投资金额和项目类型为出发点，同时考虑到年度增长率、项目批复影响因素、物资单价浮动、设备使用状态等要素，对小类采购量进行回归分析，剔除不显著的影响因素，得到多因素的回归系数；
- 4) 最后通过得到的多元回归模型以及未来一段时间内相关因素的取值预测各类物资采购需求量。根据近三年所有项目的采购订单信息，包括项目类型、项目金额、小类金额占比、电压等级等，对项目进行聚类^[3] (聚类整体分布情况见表 1)。



Figure 3. Up-bottom demand forecasting model framework for power materials
图 3. 电力物资自上而下需求预测模型框架

Table 1. Overall distribution of item clustering

表 1. 项目聚类整体分布情况

第 1 类		第 2 类		第 3 类	
电缆保护管	97.66%	10 kV 变压器	73.94%	电力电缆	86.00%
电力电缆	1.90%	架空绝缘导线	12.97%	电缆保护管	6.78%
普通光缆	0.10%	锥形水泥杆	8.27%	柱上断路器	1.38%
电缆附件--其他	0.09%	钢管杆(桩)	1.19%	锥形水泥杆	0.93%
锥形水泥杆	0.07%	接地短路故障指示器	0.65%	架空绝缘导线	0.60%
柱上断路器	0.07%	柱上断路器	0.48%	高压开关柜	0.56%
低压电力电缆	0.05%	高压计量柜	0.45%	电缆附件--其他	0.53%
光纤配线架(ODF)	0.03%	环网柜	0.38%	配电终端	0.36%
交流避雷器	0.01%	电力电缆	0.33%	线路保护	0.27%
钢管杆(桩)	0.01%	交流避雷器	0.33%	环网柜	0.25%
第 4 类		第 5 类		第 6 类	
环网柜	68.12%	锥形水泥杆	57.03%	架空绝缘导线	85.45%
电力电缆	8.45%	钢管杆(桩)	22.28%	锥形水泥杆	5.24%
配电终端	7.79%	架空绝缘导线	8.35%	钢管杆(桩)	3.57%
交流电源系统	3.59%	接地短路故障指示器	3.38%	10kV 变压器	1.68%
变电在线监测装置	2.41%	交流避雷器	2.63%	柱上断路器	1.34%
箱式变电站	1.82%	电力电缆	1.45%	低压电力电缆	0.63%
压力变送器	1.27%	高压计量柜	1.19%	交流避雷器	0.56%
低压屏(柜)、箱	1.20%	10kV 变压器	1.17%	高压计量柜	0.44%
其他	0.94%	柱上断路器	1.08%	电力电缆	0.33%
自动化系统及设备	0.83%	低压电力电缆	0.67%	接地短路故障指示器	0.29%
第 7 类		第 8 类		第 9 类	
架空绝缘导线	43.52%	高压计量柜	93.60%	箱式变电站	14.97%
锥形水泥杆	18.81%	导、地线-其他	3.76%	接地短路故障指示器	11.35%
10kV 变压器	15.92%	交流避雷器	1.24%	高压开关柜	7.63%
钢管杆(桩)	7.96%	其他	0.63%	柱上断路器	6.79%
柱上断路器	6.38%	等径水泥杆	0.28%	低压电力电缆	6.76%
电力电缆	2.37%	低压电力电缆	0.28%	配电终端	5.41%
接地短路故障指示器	1.21%	锥形水泥杆	0.21%	三相智能电能表	4.89%
其他	1.16%	架空绝缘导线	0.00%	线路保护	4.36%
交流避雷器	1.07%	10kV 变压器	0.00%	电力电缆	4.19%
低压电力电缆	0.41%	钢管杆(桩)	0.00%	直流电源系统	3.60%
第 10 类		第 11 类			
低压开关柜	49.65%	导、地线-其他	66.00%		
环网柜	15.17%	其他	30.35%		
低压电容器柜	14.41%	高压计量柜	2.46%		
封闭绝缘母线	5.69%	电缆保护管	0.42%		
电力电缆	2.63%	等径水泥杆	0.39%		
10kV 变压器	2.29%	锥形水泥杆	0.38%		
配电终端	1.83%				
变电在线监测装置	1.37%				
压力变送器	1.30%				
交流电源系统	0.73%				

3. 预测效果分析

3.1. 小类层面

本次分析一共涉及 90 个小类，从小类数量的角度进行统计，协议库存单批预测与批次提报偏差等级分布情况如表 2。

Table 2. Distribution of sub-category prediction effects

表 2. 小类预测效果分布

置信等级	小类数目	小类数目占比
一	57	63.33%
二	20	22.22%
三	7	7.78%
四	6	6.67%
总计	90	100.00%

3.2. 单个物料层面

本次分析共有 159 种协议库存，从数量的角度进行统计，批次提报与协议库存单批预测的偏差等级分布情况如表 3。库存单批预测效果较好(预测置信等级为一、二级)的单个物料占比达到了 41%。

Table 3. Predictive effect of single material

表 3. 单个物料预测效果

置信等级	物料数目	小类所含物料数目占比
一	37	23.27%
二	28	17.61%
三	37	23.27%
四	57	35.85%
总计	159	100.00%

3.3. 总金额层面

在分析的项目范围内，2018 年 1 月到 2018 年 6 月实际采购总金额为 8.08 亿元，按物料维度预测时，预测采购金额为 7.72 亿元，预测偏差百分比为 4.4%。因此，整体采购需求预测精度较优。

4. 前景展望

4.1. 机器学习

在提供发现埋藏数据深层模式的能力上，机器学习有着潜在的能力，使得应用程序更加的强大并且更能响应用户的需求。机器学习方法中如支持向量机预测模型，它没有太多的先决条件，完全依靠数据建模，挖掘数据之间的关系。原则上给予足够多的时间，支持向量机预测能刻画任意的函数关系。这种预测方法有较好的学习能力以及适应能力，预测精度能随着实际应用逐步提高[4]。

4.2. 大计划体系

通过项目全程可视化平台的运作，实现多项目管理，多方主体协同录入和查询信息，达到需求计划

和供应计划信息的可视化协同化，并形成实时执行分析与监控预警结果，便于计划人员实时掌控预测效果和批次执行效果。最终打造以计划为核心的计划 - 执行 - 分析 - 预警 - 监控体系，构建电力物资智慧供应链，实现业务全程决策智能化、协同化、可视化的目标[5]。

参考文献

- [1] 王灿. 协议库存计划申报执行[J]. 低碳世界, 2017(18): 65-66.
- [2] 尉欣. 浅议协议库存物资计划预测管理[J]. 企业改革与管理, 2015(21): 216-217..
- [3] 周俊超, 彭娟, 张靓, 田传波. 基于回归分析法的农配网协议库存需求预测[J]. 电信科学, 2017(S1): 184-189.
- [4] 国家电网公司. 国家电网公司物力集约化管理实践与创新[M]. 北京: 中国电力出版社, 2015.
- [5] 魏华, 王家冕. 供应链管理模式下的协议库存物资计划预测管理探讨[J]. 中国电力教育, 2014(9): 210-212.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2160-7311, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>
期刊邮箱: mm@hanspub.org