

Design and Implementation of Software License Demand Forecasting Model Based on Probability Statistics

Zhigang Li, Xue Feng

Information Technology Center of CNOOC, Beijing
Email: lizhg5@cnooc.com.cn

Received: Feb. 9th, 2019; accepted: Feb. 21st, 2019; published: Feb. 28th, 2019

Abstract

To master the real demand of the license of scientific research software of the group company and improve its utilization, we discover different rules of peak-valley cycles in the use of software license through analyzing mass monitoring data from branch companies. We used the shared method to design and establish a license application center, which centralized management and realized sharing of enterprise software within the group company, reduced the number of licenses purchased by the group company, and saved software investment costs. This study proposes a data-driven license demand forecasting method, which provides a scientific basis for enterprise software planning.

Keywords

Probability Statistics, Data Acquisition, Demand Prediction, Remote Share

基于概率统计的软件许可证需求数量预测模型设计与实现

李志刚, 冯 雪

中国海洋石油集团有限公司信息技术中心, 北京
Email: lizhg5@cnooc.com.cn

收稿日期: 2019年2月9日; 录用日期: 2019年2月21日; 发布日期: 2019年2月28日

摘 要

为降低集团公司科研软件采购成本, 提高软件许可证利用率, 通过对许可证使用数据的统计, 利用概率

统计的方法构建许可证需求预测模型。根据各分公司软件需求时间各异的规律, 利用削峰平谷方法设计并建立集团公司科研软件许可证共享中心, 实现软件集中管理与共享应用, 减少集团公司许可证采购数量, 节约软件投资成本, 该研究提出基于概率统计的许可证需求预测方法, 为企业版软件许可证采购数量提供依据。

关键词

概率统计, 数据采集, 需求预测, 远程共享

Copyright © 2019 by author(s) 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

随着海洋石油工业的快速发展, 科研软硬件在勘探开发研究中发挥重要的作用。油气综合研究中多学科、多专业协同, 为油气储量预测与决策提供技术保障[1] [2]。在石油勘探领域中有大量的解释专业软件被频繁使用, 但由于软件许可证数量有限, 导致日常科研与生产工作不能正常进行[3]。科研软件许可证昂贵, 提高其利用率是企业降本增效的有效方式。宋付英等从大型企业软件统一管理角度探索了软件共享对企业提质增效的作用[4]。常大明利用远程接入技术解决单机加密软件共享的问题, 将专业软件中 C/S 架构通过 SUN SGD 和 CITIRX 虚拟化技术实现 B/S 应用实现软件随时随地应用[5]。魏锋等设计了专业软件远程共享 100 ms 圈内的应用模式, 有效解决跨国石油公司专业软件软硬件资源共享应用上存在的问题[6]。以往软件共享研究主要通过架构设计与改变软件的应用模式等实现共享应用, 然而, 缺少对于软件许可证集中监测和许可证使用数据的分析。集团公司各所属单位软件许可证独立使用, 只能通过任务量和技术人员数量来估算需求。随着各单位科研任务和人数不断增加, 许可证总数成倍数增长, 科研软件价格昂贵, 为满足科研生产需要, 公司投入巨额资金购买许可证。近年来, 国际油价持续波动, 上游业务受到严重影响[7] [8], 持续高投入面临挑战。

本文在获取许可证海量监测数据基础上, 利用统计分析方法, 挖掘各单位软件许可证使用峰谷周期规律, 建立集团公司许可证共享中心, 提出基于数据驱动的许可证需求预测方法, 为公司科研软件应用规划提供科学依据。

2. 需求预测模型设计

2.1. 数据采集与处理

2.1.1. 跨平台许可证使用数据采集

数据采集是数据监测的基础[9]。科研软件种类多, 许可证服务运行在 linux、windows 等异构系统下, 国内外软件厂商采用不同的许可证管理模式, 包括 Flexlm、SafeNet、底层 IO 请求等管理模式, 针对不同软件许可证数据采用专有方法, 并将采集数据标准格式与非标准格式重新结构化输出到数据库中, 进行数据归一化处理, 解决了跨平台、跨厂商科研软件许可证数据采集的难题。

2.1.2. 数据处理与可视化

海量数据展示基于 ASP.NET 2.0 进行开发, Python 语言编程实现, 底层数据库为 SQL SERVER 数据

库, 利用许可证实时数据缓存技术实现海量数据存储。数据展示为使用者提供便利, 可视化技术出现在科学计算可视化(Visualization in Scientific Computing), 指利用计算机图形学、计算机图像处理、计算机信号处理等方法对数据、信息、知识的内在结构进行表达。在数据的种类与范围与许可证使用数据存在差异。许可证监测数据从多个维度迅速直观展示许可证使用情况需求, 研发许可证监测可视化技术, 集合多个数据图表, 支持多表关联、追加合并, 包括使用频率、峰谷周期、购买数与使用数合理性分析等, 实现跨系统、跨专业许可证应用数据多角度立体展示功能。

功能实现主要方法:

```
public void createReport(String software, String
startDate, String endDate) {
    List<LicenseInfo> licenseNameList =
getLicenseNameList(software);
    View report = new View();
    //创建汇总页
    createSummary();
    //遍历软件许可 List
    for(LicenseInfo li:licenseNameList) {
        //创建分项页
        createItem();
        //绘制图表
        drawChart(report);
    }
}
public void drawChart(View report) {
    //创建图表
    ChartShape chart = report.addChart();
    //设置数据系列
    chart.setSeriesName();
        chart.setSeriesYValueFormula();
        chart.setCategoryFormula();
    //设置图例
    chart.setLegend();
    //设置坐标轴范围
    chart.setScaleValueRange();
}
```

组合显示多样化: 显示多软件、多模块在各个单位的使用数、购买数、最大值, 并自动集成绘制趋势图, 如图 1 所示。

指标统计自动化: 生成软件使用汇总报告, 统计软件的整体使用情况, 包括该软件所有模块在所有单位的使用峰值及使用率, 并自动绘制使用峰值统计图, 如图 2 所示。

2.2. 数据整合与分析

在采集的各个单位许可证使用数据的基础上, 通过可视化数据的分析, 发现了科研软件使用的峰谷规律。

以综合解释软件 GeoFrame 某模块的使用情况为例, 集团公司采购许可证总数 120 个, 分配给各单位用于科研生产, 各单位峰值累加为 99 个, 许可证的使用率为 82.5%, 软件使用饱满。

然而, 将各单位监测数据叠加后, 叠加峰值为 79 个, 许可证使用率为 65.2%。

各单位应用峰谷不重叠, 存在错峰现象, 使软件共享应用成为可能。如图 3 所示。

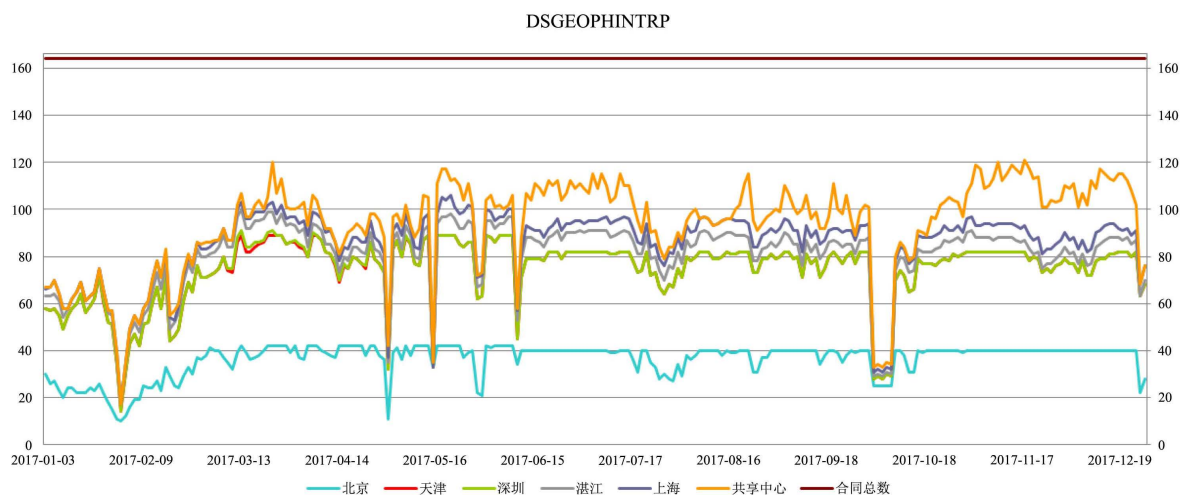


Figure 1. Multi-unit usage of the same module
图 1. 同一模块多单位使用情况

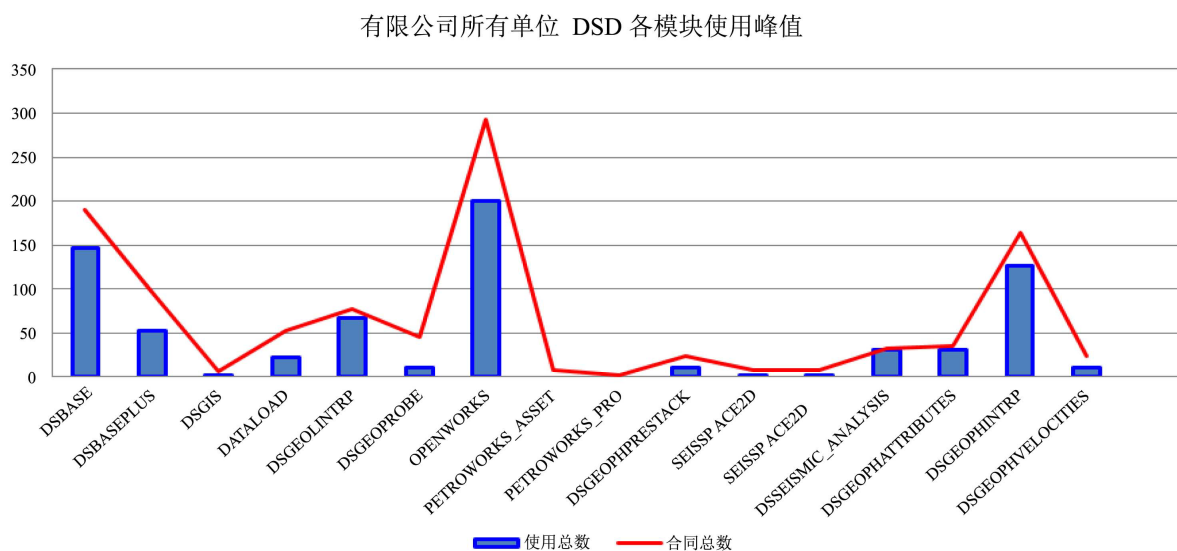


Figure 2. Peak of the same module
图 2. 同一模块峰值

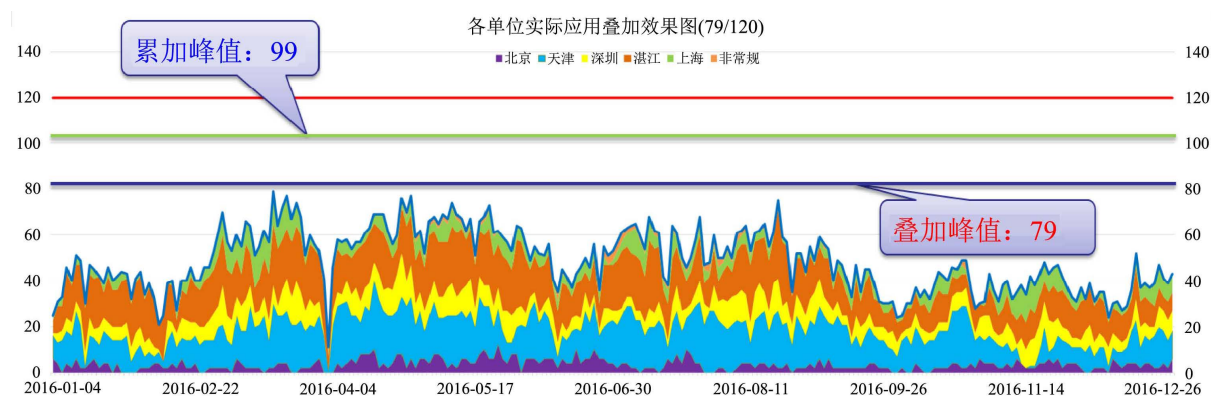


Figure 3. A Geo Frame module stacking peak
图 3. 软件 Geo Frame 某模块叠加峰值

2.3. 基于滑动平均算法的许可证需求预测

概率统计在经济社会的各个方面应用广泛, 为解决现实世界中的问题提供研究思路[10] [11]。通过对集团公司许可证使用数据监测, 利用滑动平均算法预测许可证需求数量趋势, 结合科研任务量和人员数量, 为集团公司许可证采购与制定许可证共享分配策略提供依据。滑动平均的计算公式如下:

$$F_t = (A_{t-1} + A_{t-2} + A_{t-3} + \dots + A_{t-n}) / n,$$

- F_t 代表预测值;
- n 代表移动平均的时期个数;
- A_{t-1} 代表前期实际值;
- A_{t-2} , A_{t-3} 和 A_{t-n} 分别表示前两期、前三期直至前 n 期的实际值。

3. 应用实例

以北京研究中心软件某模块 2016 年 1 月至 2017 年 6 月许可证使用情况数据为例。滑动平均拟合曲线反映用户使用情况。红色曲线代表滑动平均。如图 4 所示。

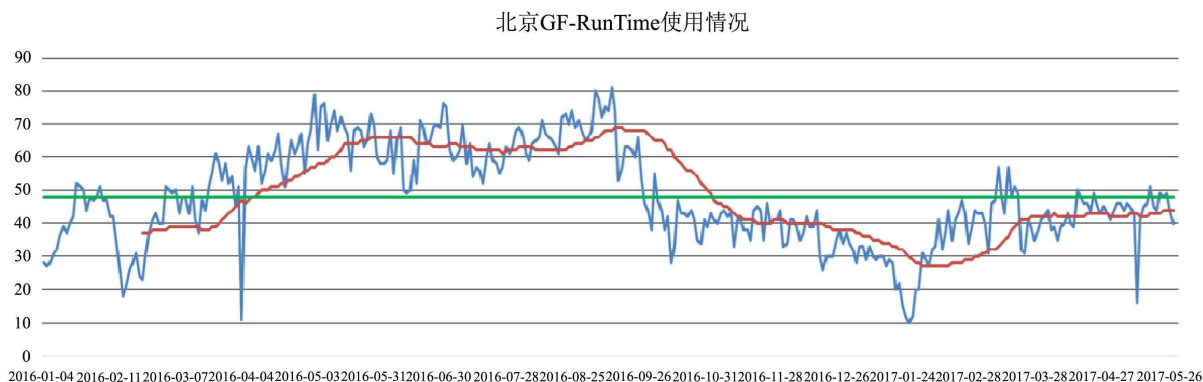


Figure 4. Sliding average fitting license use schematic map

图 4. 滑动平均拟合许可证使用示意图

绿色代表需求的均值。从概率统计规律看, 这条拟合曲线离散数据服从正态分布, 如图 5 所示。利用正态分布理论, 在均值上递减一个标准差 ($\mu - \sigma$), 可满足约 30% 用户应用需求。

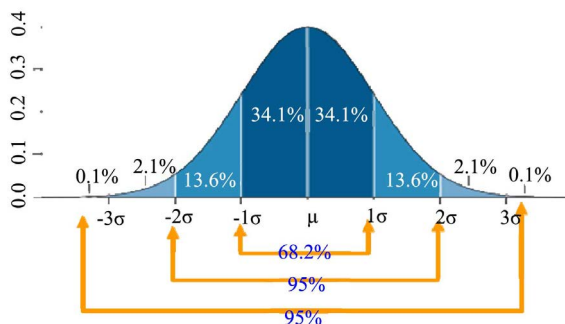


Figure 5. Distribution range of license demand

图 5. 许可证需求分布区间

标准差公式:
$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad E(\bar{X}) = \mu$$

依据正态分布理论, 利用标准差公式计算出许可证保留数为 35 个(紫色直线), 经测算能支持 68% 用户需求, 为制定集团公司购买许可证数量提供数据支撑。如图 6 所示。

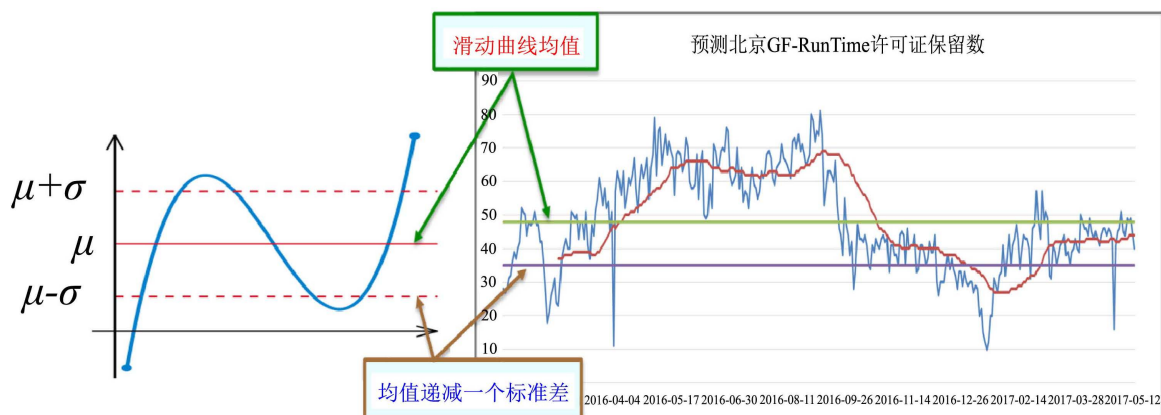


Figure 6. Rational section of license purchase
图 6. 许可证采购合理区间

1) 通过对采集 8000 多万条监测数据分析和设计科研软件许可证需求预测模型, 为集团公司采购许可证提供决策依据;

2) 许可证共享中心为 1800 多名科研人员服务, 该共享应用模式相比传统模式一次性节约采购成本 25%, 经济效益显著。

4. 结束语

专业软件许可证跨操作系统、跨软件厂商、跨区域共享应用模式, 突破传统独立使用的壁垒, 提出的基于滑动平均算法的许可证需求数量预测方法为采购许可证数量提供详实数据支撑, 发现的用户使用许可证数量趋势符合泊松分布规律为公司制定软件采购规划提供科学依据。集团公司范围的科研软件许可证共享应用新模式为公司海外科研软件共享应用发挥示范作用。

参考文献

- [1] 陈明强, 李允, 黄纯国, 等. 油气勘探开发综合信息系统集成方案的设计与实现[J]. 西南石油大学学报(自然科学版), 1998, 20(1): 25-28.
- [2] 李慧, 李志刚, 王茹, 刘锐, 孙超. 云技术在油气田勘探开发系统建设中的应用研究[J]. 网络安全技术与应用, 2018(5): 85-86.
- [3] 杨显峰, 毛承国, 陈翔, 等. 软件远程共享管理平台的设计与实现[J]. 计算机时代, 2016(9): 42-44.
- [4] 宋付英, 陈新荣, 蒋多元, 等. 企业软件资源共享管理策略研究[J]. 中国管理信息化, 2018(11): 51-52.
- [5] 常大明. 基于远程共享技术的勘探开发专业软件远程共享平台[J]. 信息与电脑(理论版), 2017(7): 42-43.
- [6] 魏锋, 刘瑾, 张超林, 等. 跨国公司石油专业软件共享系统的设计与应用[J]. 石油化工自动化, 2016, 52(3): 50-54.
- [7] 冯启海, 卫永刚, 郑德鹏, 等. 国内外石油企业应对低油价的做法及启示[J]. 国际石油经济, 2016, 24(7): 39-43.
- [8] 李志刚, 王俊琴. 基于 Studio 的 Petrel 数据共享环境建设与应用[J]. 网络安全技术与应用, 2016(2): 65-66.
- [9] Kandadai, V., Sridharan, M., Parvathy, S.M. and Pitchaimuthu, R. (2017) A Comprehensive Embedded Solution for Data Acquisition and Communication Using FPGA. *Journal of Applied Research and Technology*, 15, 45-53. <https://doi.org/10.1016/j.jart.2016.12.002>
- [10] 陈皓. 概率统计在彩票中的应用探析[J]. 企业科技与发展, 2018(2): 193-194.
- [11] 张森源. 概率统计在实际生活中的应用分析[J]. 经贸实践, 2018(2): 32.

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：2161-8801，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：csa@hanspub.org