

出行产品未来14个月销量预测 说明文档

团队：小月亮

2017年5月2日

CONTENTS
目录

1

团队介绍

2

赛题分析

3

算法流程

4

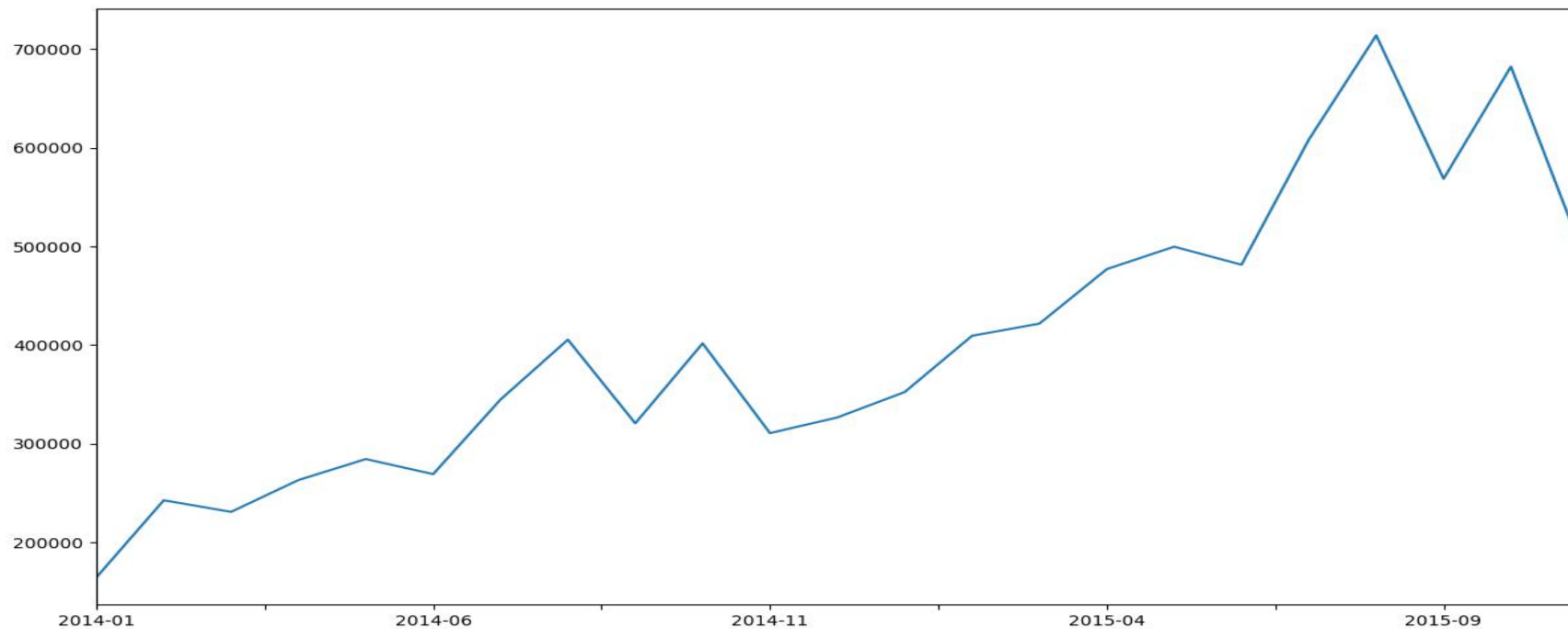
问题总结与致谢

► 1. 团队介绍

姓名：张月鹏
年级：研究生三年级
专业：大连理工大学数学科学学院
[邮箱：525640678@qq.com](mailto:525640678@qq.com)
手机：18340813410
技能：Python 机器学习

姓名：杨凡
年级：研究生二年级
专业：大连理工大学经管学院
[邮箱：1498288190@qq.com](mailto:1498288190@qq.com)
手机：18041155347
爱好：机器学习 Java 自然语言处理

► 2. 赛题分析



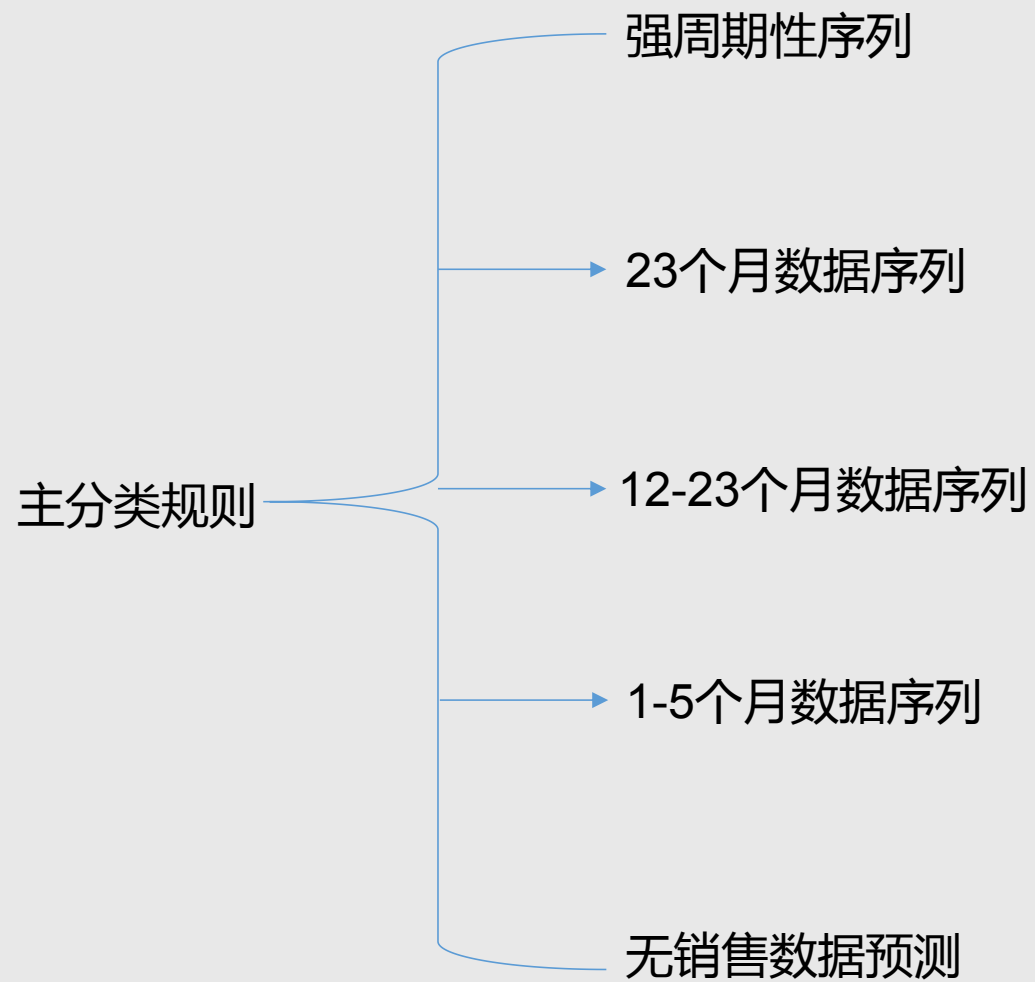
► 总体特点：

- 周期性明显
- 数据缺失值多
- 整体有增长趋势
- 要求预测时间范围长

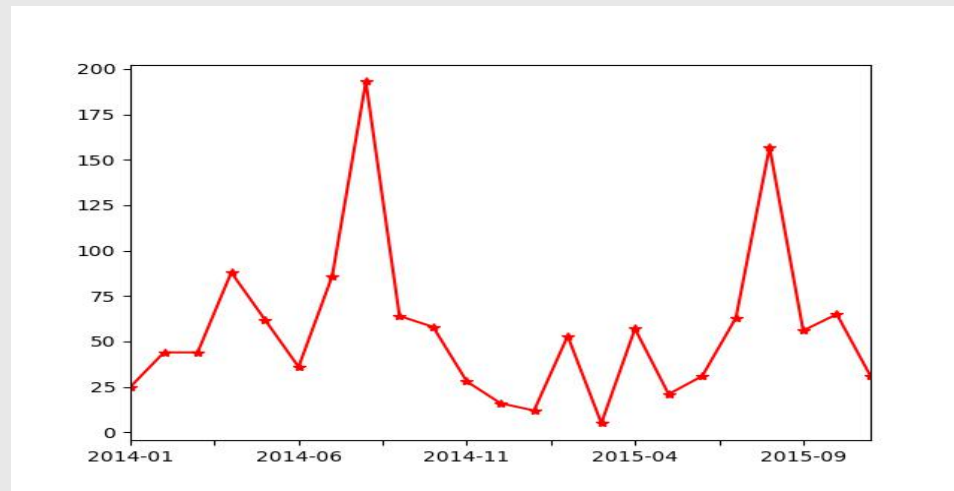
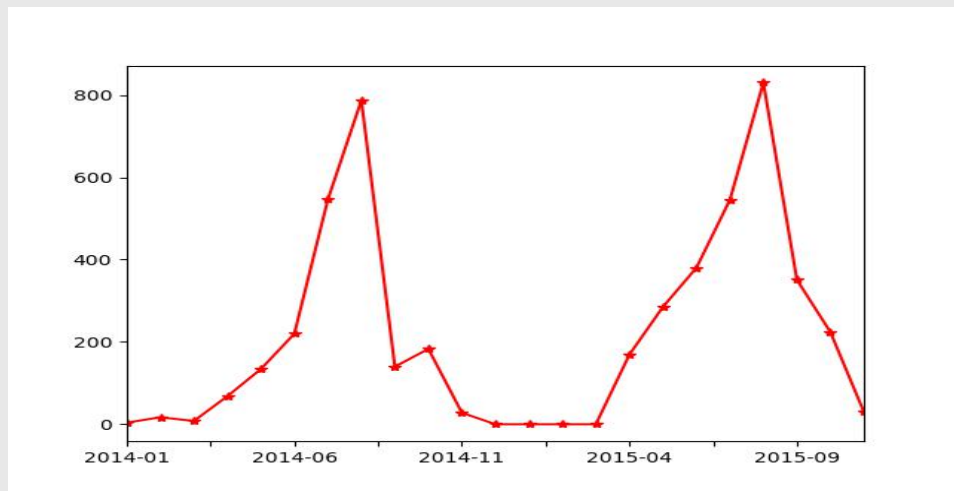
▶ 3. 算法流程

- ◆ 我们认为后12个月的数据最能反映未来14个月的基本变化情况，然而由于给定数据含有大量的缺失值，这促使我们将数据进行分类，分别预测。
- ◆ 采用逐步剥离的方式寻找特征，并根据首次具有销量的时间将数据分为23个月数据、12-23个月数据、1-12个月数据、无销量这四个类别。其中23个月数据表示首次具有销量的月份为给定数据的第一个月。
- ◆ 考虑其较小的数据量和较大的随机性，我们采用统计和规则的方式从时间序列本身寻找规律进行预测。

▶ 3. 算法流程

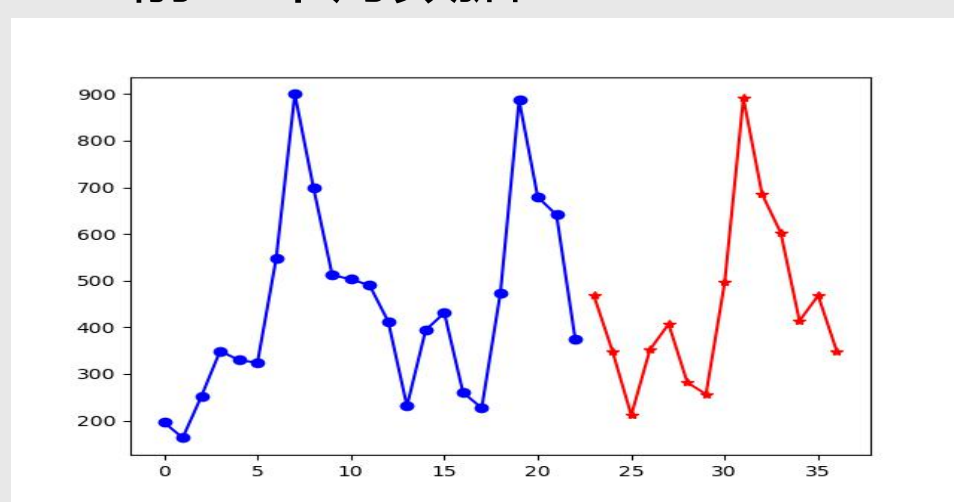
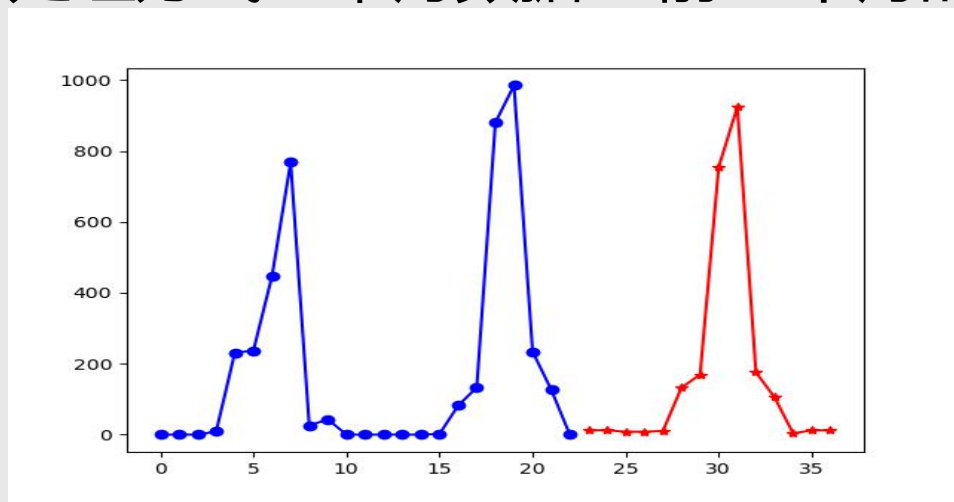


具有强周期性的时间序列



特征：数据量中最高的两个值时间相差12个月

处理方式：单月数据 = 前12个月的数据 * 0.7 + 前24个月数据 * 0.3



► 23个月的时间序列

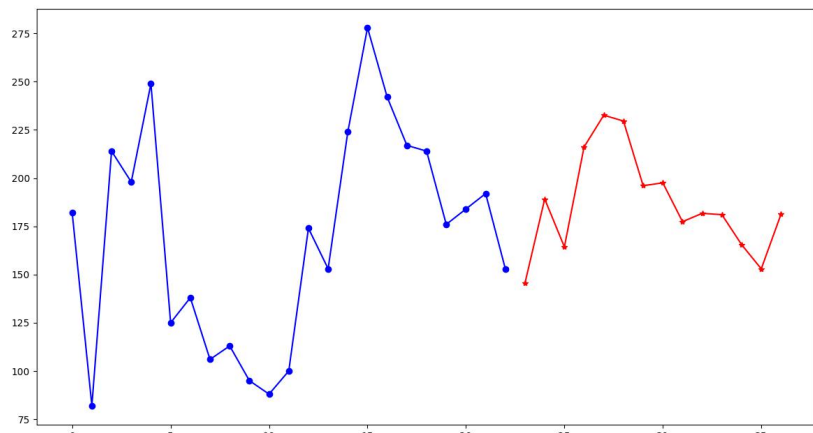
针对23个月的时间序列预测，将时间序列分解为：

时间序列 = 基本项 * 基本项系数 + 周期项 * 周期项系数 + 趋势项 * 趋势项系数 + 随机噪声

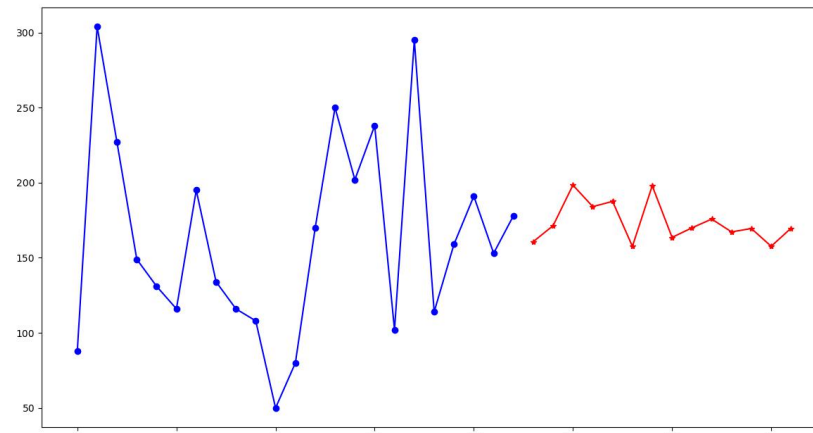
对于一个产品的预测，我们使用统计规则判断该产品的周期项和趋势项。

- 基础项使用前12个月的数据。
 - 前6个月的平均值 * 系数1 + 前12个月的平均值 * 系数2。
- 周期性判断：
 - 使用前11个月和后11个月的相关系数来计算周期性，相关系数越大认为周期性越高，周期项系数越大。
- 趋势项判断：
 - 对数据进行两次移动平均进行趋势判断，如果其后期增长时间较长并且基本方式求出的均值结果较小，则认为其有增长趋势，然后利用斜率的中位数来对增长趋势进行刻画。

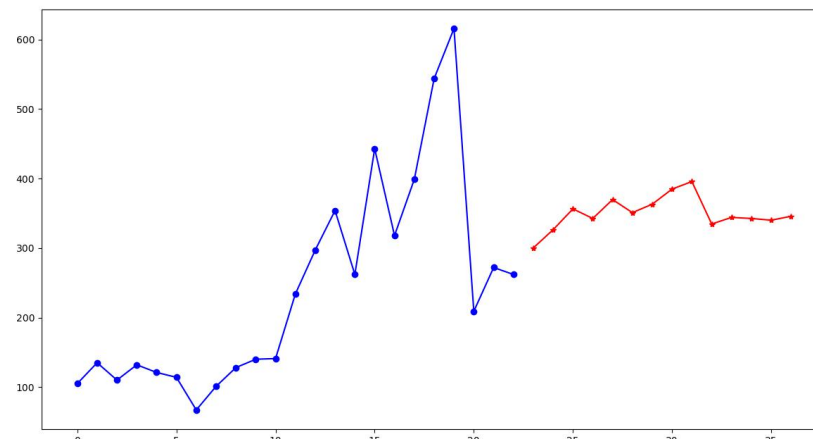
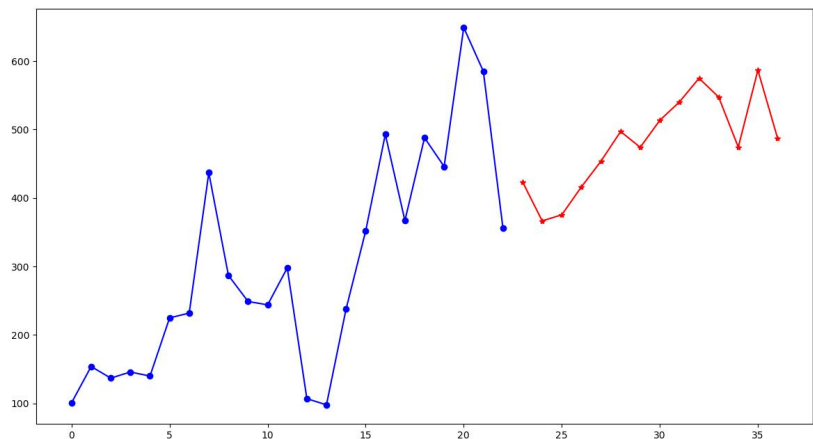
23个月的时间序列



图例1、2



图例3、4



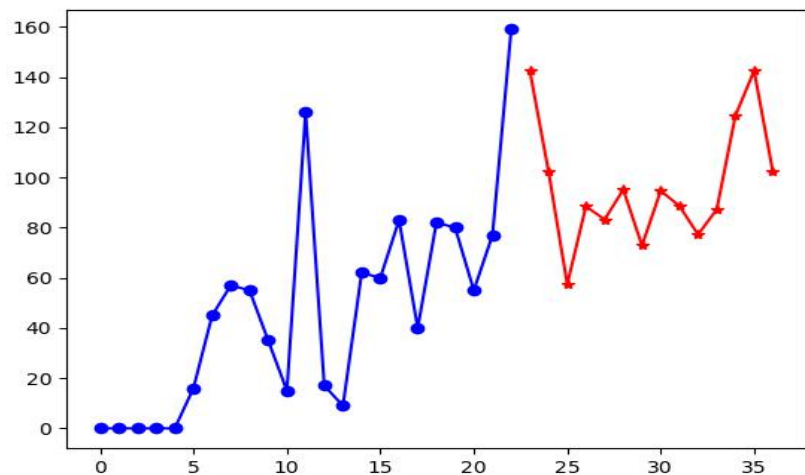
► 12-23个月的时间序列

由于12-23个月的数据量较少，不易判断产品的周期性和整体增长趋势，故使用新的规则进行处理。

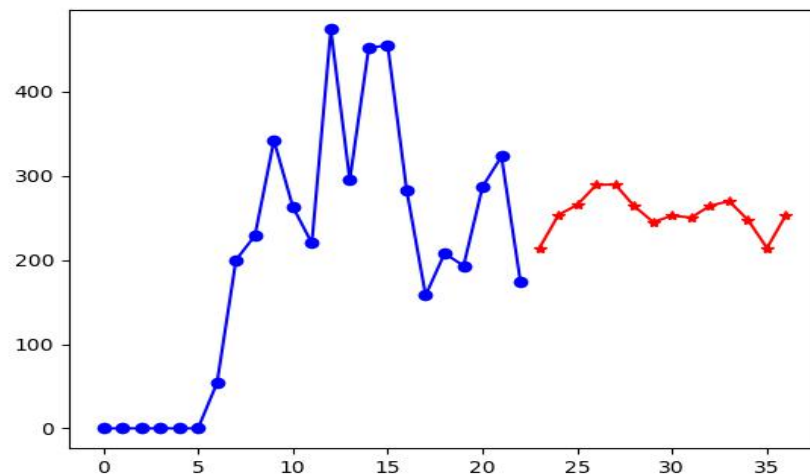
我们分析了地区三与产品的销量具有一定的相关性，使用地区三来判断产品的周期性；使用后12个月的数据判断具有增长趋势的序列，将具有明显增长趋势的序列剥离出来单独处理。

- 周期性判断：
 - 规则：使用地区三的销量数据来判断产品周期性，若某一产品的地区三表现出具有强周期，则认为该产品的具有周期性的概率更大。
 - 处理方法：若该产品的地区三为强周期序列，周期项系数增大为0.45，否则为0.15。
- 增长趋势：
 - 规则：后4个月的均值 > 4-8个月的均值 > 8-12月的均值。
 - 处理方法：在周期项上乘以增长系数，系数递减（1，0.8，0.6）。

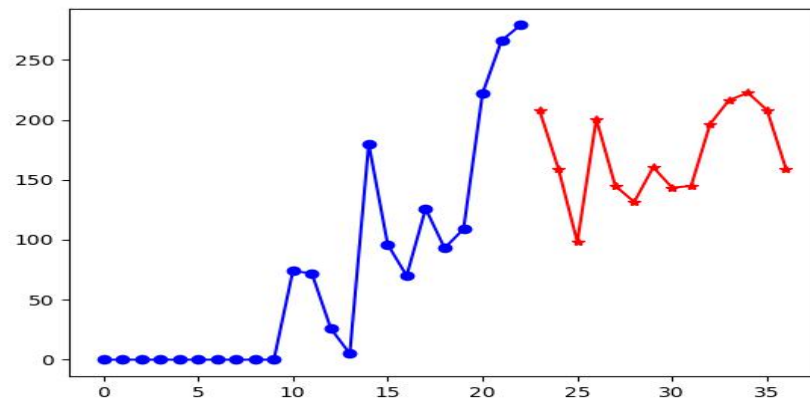
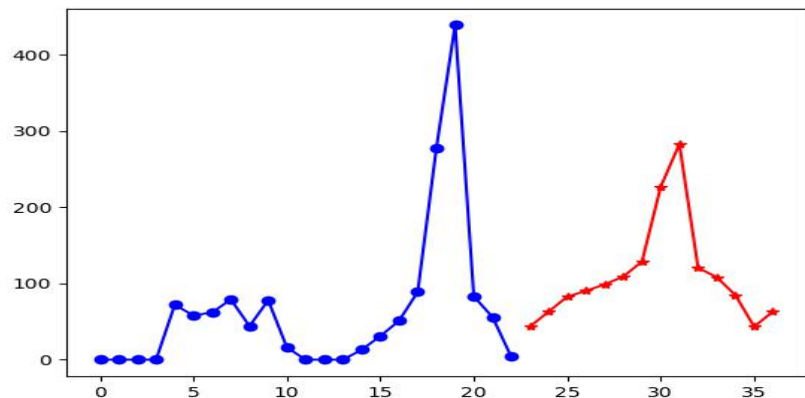
► 12-23个月的时间序列



图例1、2

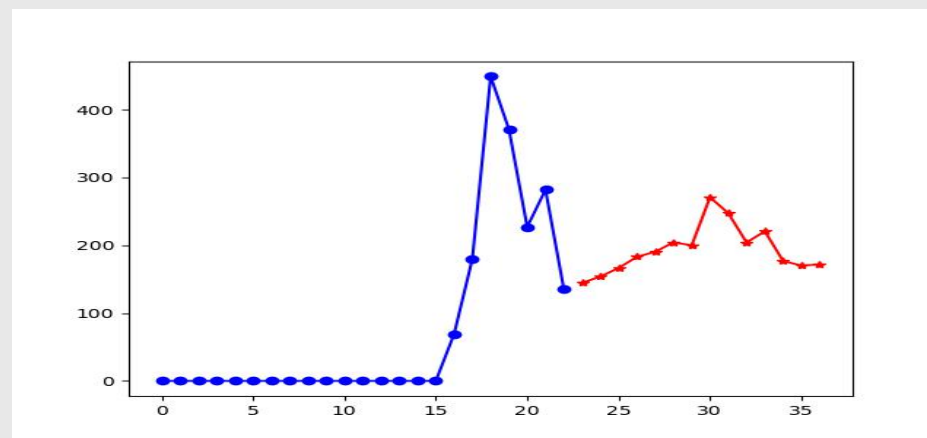
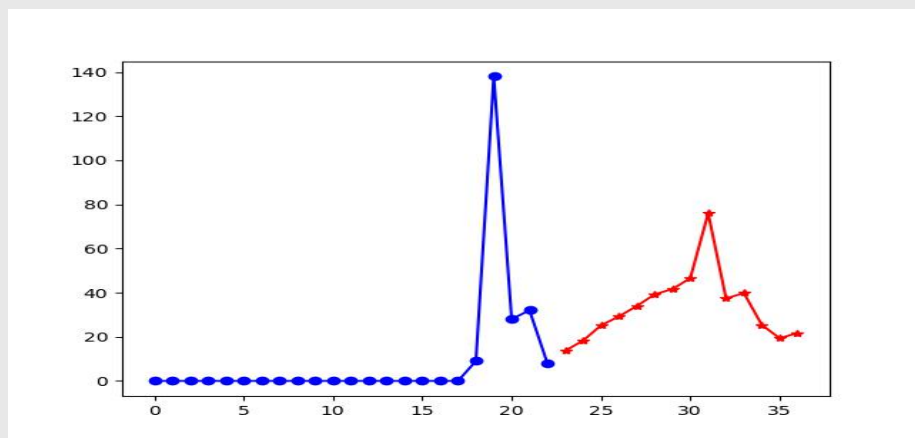
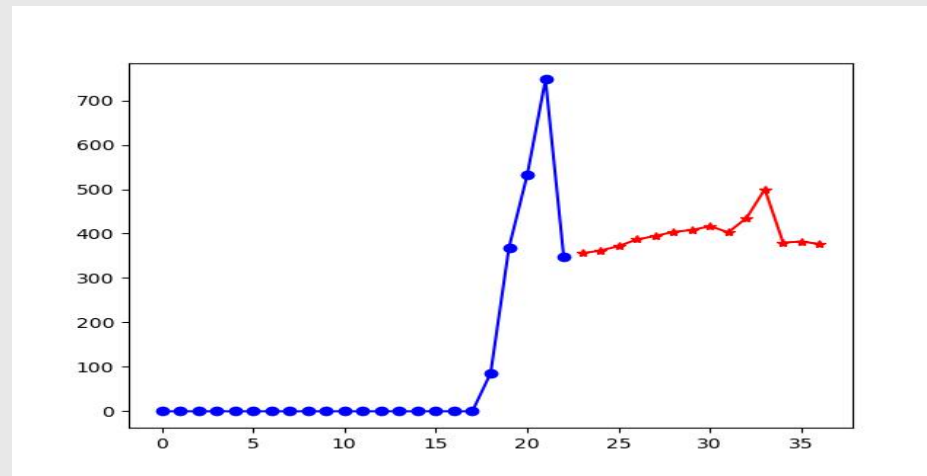
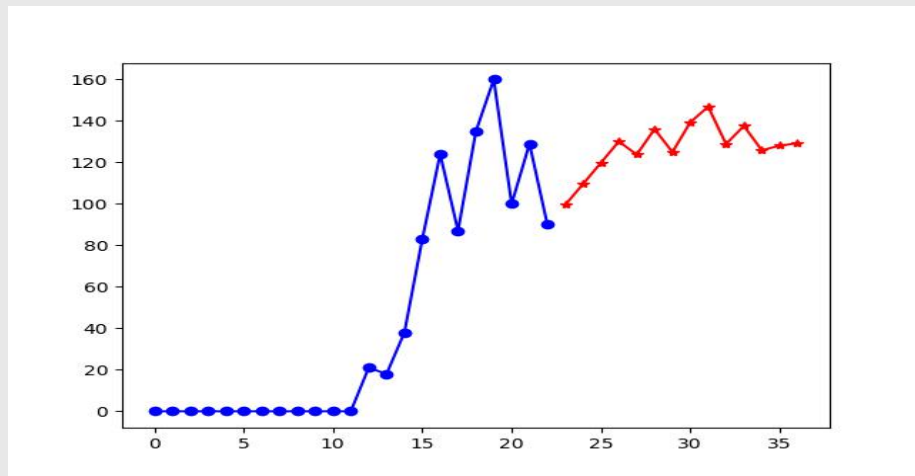


图例3、4



► 1-12个月的时间序列

由于其销量月份比较少难以有效学习销量，而早期销量又不够稳定，所以算法我们比较多的依赖的后期销量以及voters修正处理。



▶ 无销量预测

- 算法在预测没有销量数据时，我们默认从合作时间和开始时间（我们会对时间的合理性进行判断）的最小值开始处理产生销量。
- 利用地区三的趋势来估计产品的趋势
- 重点利用voters来预测平均销量。

谢谢！