



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY



安泰经管学院
ANTAI COLLEGE
Economics · Management

《预测分析与方法》课程报告



上海交通大学

SHANGHAI JIAO TONG UNIVERSITY

《预测分析与方法》

Team 6 课程报告

题目：共享单车使用量的分析与预测

学 院： 安泰经济与管理学院

年 级： 大三

组 员： 王辰君 胡叶皓 邱浩原 许凌瀚 赵浚哲

指导教师： 李成璋

目录

一、研究背景与意义	2
1.1 研究背景	2
1.2 数据集介绍	2
1.3 研究意义	3
二、ETS & ARIMA Model	3
2.1 天、月级数据拟合与预测	3
2.1.1 模型选择	4
2.1.2 模型拟合与预测	5
2.2 小时级数据拟合与预测	6
2.2.1 模型选择	6
2.2.2 模型拟合与预测	8
2.2.3 周中与周末数据的预测	9
三、Prophet Model	9
3.1 天级数据拟合与预测	9
3.2 小时级数据拟合与预测	10
四、LSTM-RNN Model	11
4.1 One-step recursive forecast	12
4.1.1 天级数据的拟合和预测	12
4.1.2 小时级数据的拟合和预测	13
4.2 集成残差学习 (LSTM+Prophet)	14
4.3 Moving Window 多步预测 (LSTM)	14

一、研究背景与意义

1.1 研究背景

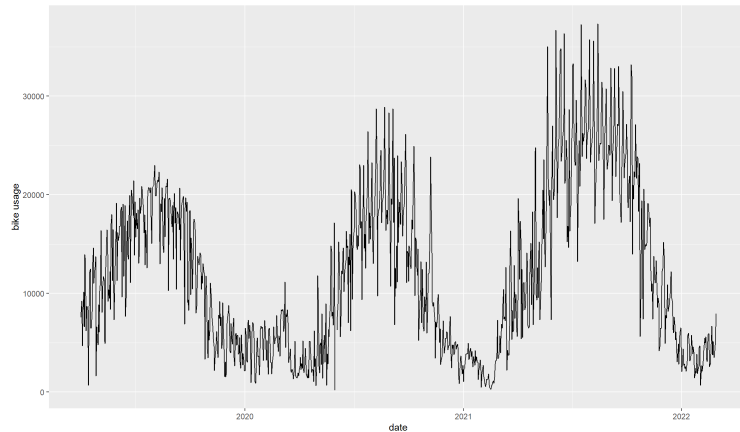
随着“节能减排”理念的不断推行，“共享”这一概念越来越为人所熟知。对于共享单车企业而言，如果能将自己的车辆在合理的时间投放到有需求的地点，那么既方便用户出行，达到节能减排的目标；又能为自身带来营业收入。相对的，如果共享单车的投放出现错配，就容易出现有些地方“供不应求”，而同时有些地方“供过于求”的现象，造成资源的闲置与浪费，对公司、客户和环境都没有好处。

因此，我们小组计划采用 ETS、ARIMA、Prophet、LSTM 神经网络等一系列模型，基于过往共享单车使用量数据，对共享单车未来的使用量进行时间序列分析与预测，帮助相关企业在未来实现共享单车的最优配置。

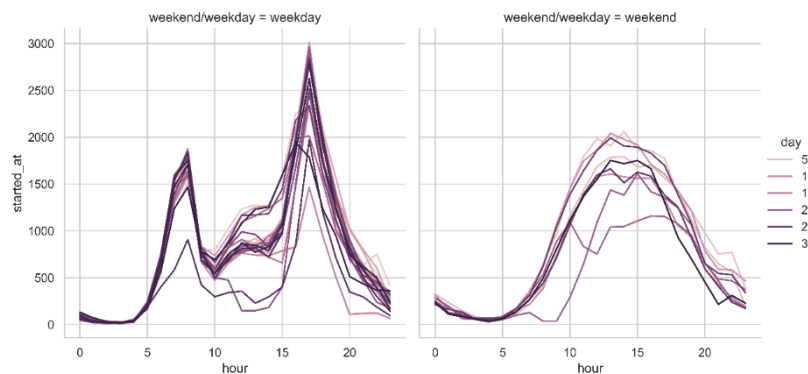
1.2 数据集介绍

我们选用的数据集为芝加哥 Cycliclic 共享单车公司 36 个月（2019 年 3 月至 2022 年 2 月）的订单数据，每一笔订单的时间都能精确到秒，方便我们后续整合出小时级、天级、月级数据进行相应分析与预测。

下图是用日级数据画的时序图：



下图是用小时级数据画的季节图，每一条线代表一天：



可以看出工作日每一天有两个自行车使用量的高峰期，分别对应早高峰与晚高峰；休息日每天只有一个高峰期，大致分布在下午。我们后续将对这些特征分别展开分析与预测。

1.3 研究意义

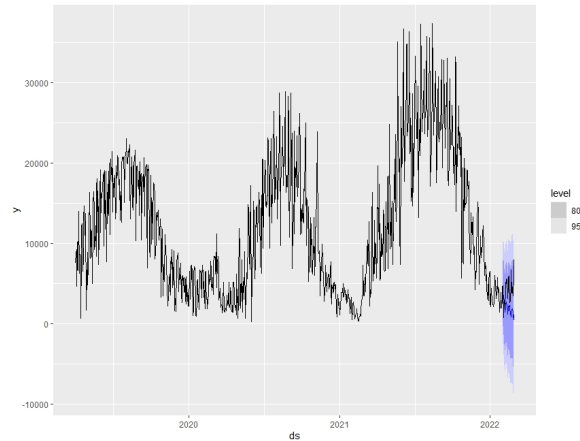
1. 小时级数据：帮助相关企业实时小范围调动共享单车
2. 天级数据：帮助相关企业提前大范围配置共享单车
3. 月级数据：帮助相关企业提前规划未来的生产产量

二、ETS & ARIMA Model

2.1 天、月级数据拟合与预测

由于 ARIMA 只能提取一个季节项的特征，因此在对具有多个季节项的时间序列数据进行拟合时，往往会出现比较大的偏差。而从天级数据的时序图中可以观测到比较明显的周周期性与年周期性，因此无法直接采用 ARIMA 模型对天级数据进行拟合。

下图展示了 ARIMA 模型对于天级数据的预测结果：

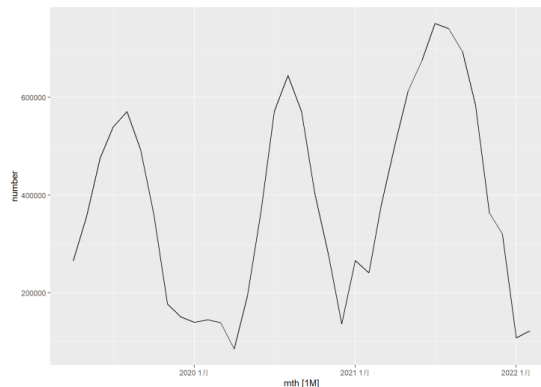


为此，我们小组考虑先将天级数据聚合成月级数据（将同一个月内的数据进行加总），消除周周期性后再用 ARIMA 模型进行拟合。

2.1.1 模型选择

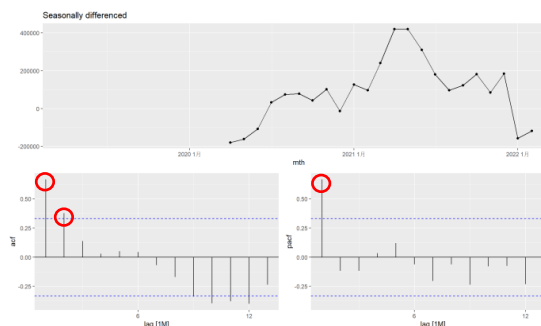
首先进行数据预处理，应用 pandas 库中的 groupby() 函数，将天级数据聚合成月级数据。然后，进行月份天数的归一化处理，将每个月的天数调整为 30 天。接着将数据导入 RStudio，应用 yearmonth() 函数将日期格式调整为年+月。

月级数据的时序图如下图所示：

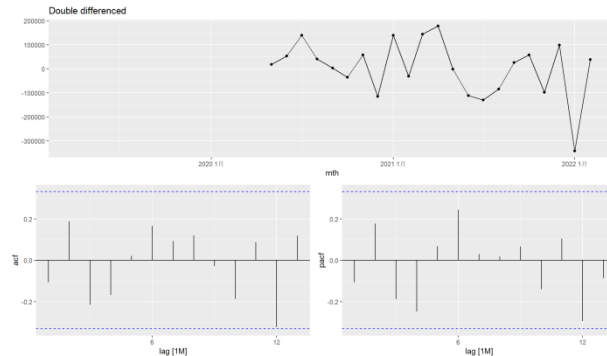


可以看出，周周期性已经完全消除，当前时序数据中还存在年周期性与一个向上的趋势性。

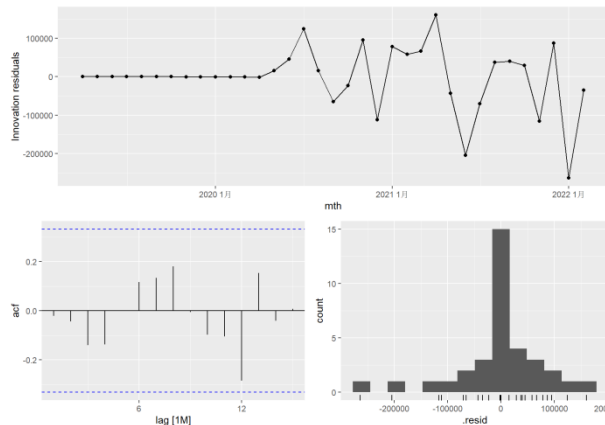
接着对月级数据进行季节性的一阶差分，如下图所示：



从 ACF 图和 PACF 图中可以看出数据可能包含非季节性的 MA (2) 和 AR (1) 成分。之后再对数据进行一阶差分，如下图所示：



下图展示了 ARIMA(1,1,2)(0,1,0)12 模型拟合的残差检验结果，可以看出所有残差都在显著性临界值内，类似于白噪声，因此我们可以将该模型用于接下来的预测。



2.1.2 模型拟合与预测

由于我们的数据集中只包含三年的数据，因此我们将前两年的数据作为训练集，第三年的数据作为测试集。

下图展示了各个模型在训练集上的拟合优度，其中包括三个 ARIMA 模型和六个 ETS 模型。可以看出，在所有的 ARIMA 模型中，我们自己对 ACF 图和偏自相关图特征中所找到的 ARIMA(1,1,2)(0,1,0)12 模型拥有最高的拟合优度；在所有 ETS 模型中，Damped multiplicative Holt-Winter's method 模型拥有最高的拟合优度。如果将这两个模型再次进行比较，ARIMA 模型的表现相对较好。

A tibble: 9 x 10

	.model <chr>	.type <chr>	ME <dbl>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>	MASE <dbl>	RMSSE <dbl>	ACF1 <dbl>
1	A_112010	Training	9026.	44216.	25700.	1.95	7.86	0.278	0.422	-0.0317
2	A_stepwise	Training	19474.	73400.	59881.	5.24	23.9	0.649	0.701	0.0440
3	A_search	Training	1738.	70512.	59376.	-8.91	28.2	0.643	0.674	-0.143
4	ETS_N_A	Training	5908.	62687.	46921.	-5.10	20.4	0.508	0.599	0.0738
5	ETS_N_M	Training	597.	54365.	45029.	-5.09	19.9	0.488	0.519	0.540
6	ETS_A_A	Training	3956.	65402.	45420.	-6.19	20.5	0.492	0.625	0.000536
7	ETS_A_M	Training	164.	53001.	43601.	-6.15	19.9	0.472	0.506	0.538
8	ETS_Ad_A	Training	4545.	65394.	45676.	-5.82	20.5	0.495	0.625	0.00353
9	ETS_Ad_M	Training	6383.	53602.	42868.	-2.16	19.1	0.464	0.512	0.531

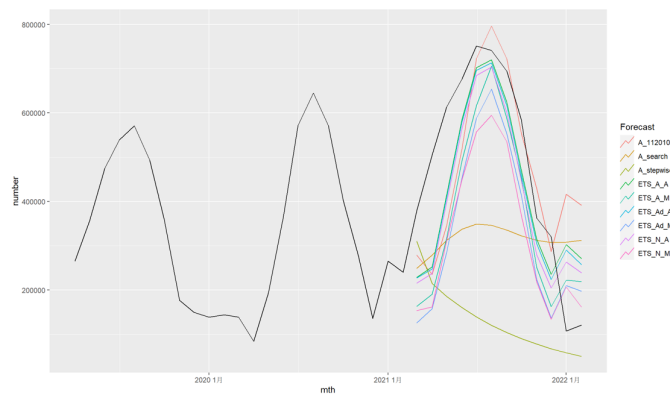
当然，在训练集上有好的表现并不能代表这个模型拥有好的预测能力，因此我们还需要关注各个模型在测试集上的表现。从下图可以看出，ARIMA(1,1,2)(0,1,0)12 模型在测试集上的表现比其他 ARIMA 模型都好，而 Holt-Winter's additive method 模型在测试集上的表现比其他 ETS 模型都好。相比之下，ETS 模型的表现比 ARIMA 更好。



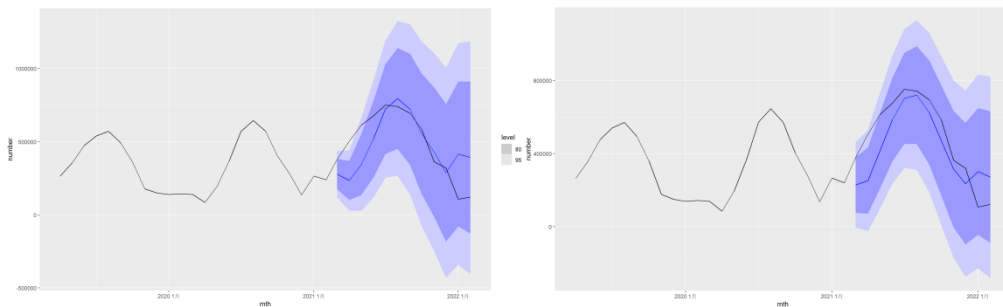
```
# A tibble: 9 x 10
```

	.model	.type	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE	ACF1
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	A_112010	Test	13295.	1.73e5	1.35e5	-31.0	58.8	1.46	1.65	0.588
2	A_search	Test	173681.	2.69e5	2.39e5	4.65	61.9	2.59	2.57	0.745
3	A_stepwise	Test	356271.	4.12e5	3.56e5	68.1	68.1	3.86	3.94	0.666
4	ETS_A_A	Test	61288.	1.37e5	1.19e5	-7.47	43.2	1.29	1.31	0.527
5	ETS_A_M	Test	123007.	1.77e5	1.58e5	12.3	43.6	1.72	1.69	0.549
6	ETS_Ad_A	Test	69214.	1.38e5	1.22e5	-4.44	42.5	1.33	1.32	0.516
7	ETS_Ad_M	Test	154883.	2.03e5	1.85e5	20.2	46.6	2.00	1.94	0.561
8	ETS_N_A	Test	84968.	1.44e5	1.31e5	1.35	41.6	1.41	1.38	0.513
9	ETS_N_M	Test	166984.	2.06e5	1.90e5	24.0	44.9	2.06	1.97	0.525

下图展示了上述各个模型的预测图。可以看出，由于没有捕捉到季节项的信息，ARIMA_search 和 ARIMA_stepwise 的预测精度都很差；其他模型的预测精度都还不错。



下面两张图分别展示了 ARIMA(1,1,2)(0,1,0)₁₂ 模型和 Holt-Winter' s additive method 模型的拟合线以及对应的置信区间。可以看出，这两个模型都捕捉到了趋势项与季节项的信息，很好地完成了对于月级数据的预测，真实值几乎完全落在 95% 置信区间之内。

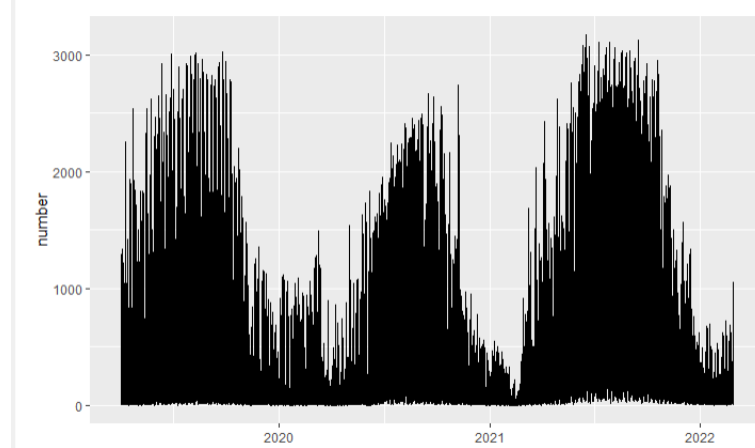


2.2 小时级数据拟合与预测

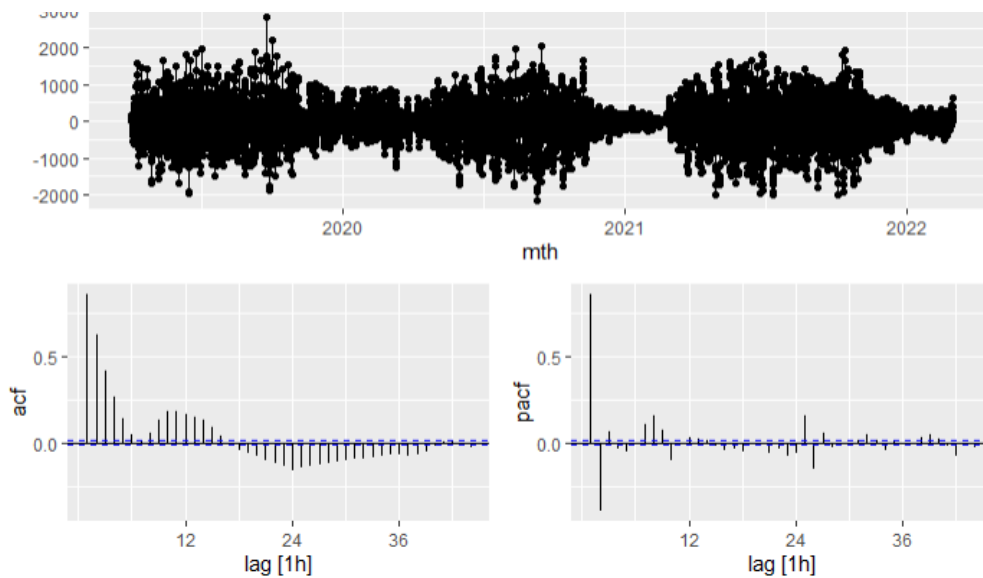
2.2.1 模型选择

采用与月级数据类似的方式进行数据预处理，应用 pandas 库中的 groupby() 函数，将所有用车数据聚合成小时级数据。然后将数据导入 RStudio，应用 as_date() 函数将日期格式调整为年+月+日+小时。接着利用 fill_gaps() 函数将没有任何用车数据的小时时点填充为 0。

小时级数据的时序图如下图所示：



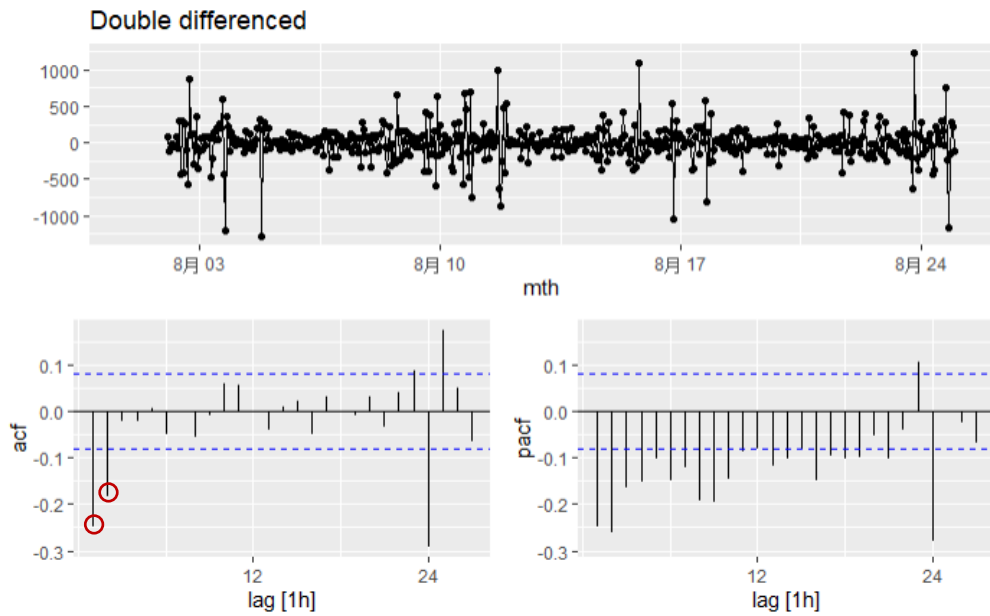
显然，小时级数据的数据极大且数据间差异性也较大，因此需要首先对数据进行差分。一阶差分后结果如下：



即使进行了差分，ACF 值和 PACF 值的分布并不能让我们得到一个对模型的合理推测，原因是相比起小时级数据的时间范围，总计 3 年的数据量过大，月度数据之间的差异导致小时级数据的波动呈现了不稳定的态势。同时，大量特殊节日的存在同样造成了巨大的差异性，对于月度数据而言，这样的差异影响不大，但对于小时级数据而言，一个节日的存在就将影响至少 24 个数据点，影响巨大。

小时级数据的研究初衷是为帮助企业实时小范围调动共享单车，因此并不需要将长期的数据作为研究的对象，因此我们考虑将研究时间跨度限定在一个月。同时，为消除节日的影响，将选取节日尽可能少的月份作为样本（考虑节日因素的模型将在后面的部分呈现）。最终，我们选定 2019 年 8 月的小时级数据作为训练集。

对数据进行季节性一阶差分后，再进行一阶差分，结果如下：



从 ACF 图和 PACF 图中可以看出数据可能包含季节性周期为 24 个数据点的 MA (2) 成分，因此选取 $ARIMA(0,1,2)(0,1,0)_{24}$ 。

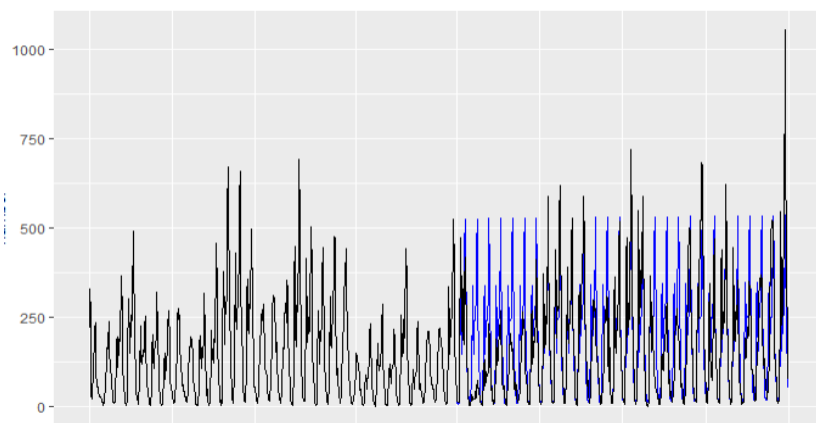
2.2.2 模型拟合与预测

首先考虑 ETS 模型，结果如下：

.model <dbl>	.type <chr>	ME <dbl>	RMSE <dbl>	MAE <dbl>	MPE <dbl>	MAPE <dbl>	MASE <dbl>	RMSSE <dbl>	ACF1 <dbl>
additive	Test	-0.46899643	44.44886	30.01349	-15.363902	46.98805	0.5798967	0.5571197	0.3820773
Damped	Test	-0.17910993	53.13737	33.91720	5.938775	37.69250	0.6553212	0.6660210	0.1264101
Holt	Test	0.07019254	58.37095	37.74003	11.974925	43.84858	0.7291828	0.7316184	0.2103654
multiplicative	Test	4.54419671	51.77300	30.58574	-11.464018	38.30584	0.5909534	0.6489201	0.2870786
SES	Test	-0.12125534	54.86065	38.62777	-15.192086	43.13034	0.7463351	0.6876206	0.4570902

所有模型的预测精度均较差，这证明了 ETS 并不是针对小时级的一个合适的模型。

其次考虑 ARIMA 模型，将训练集后的两周作为测试集，预测结果如下：

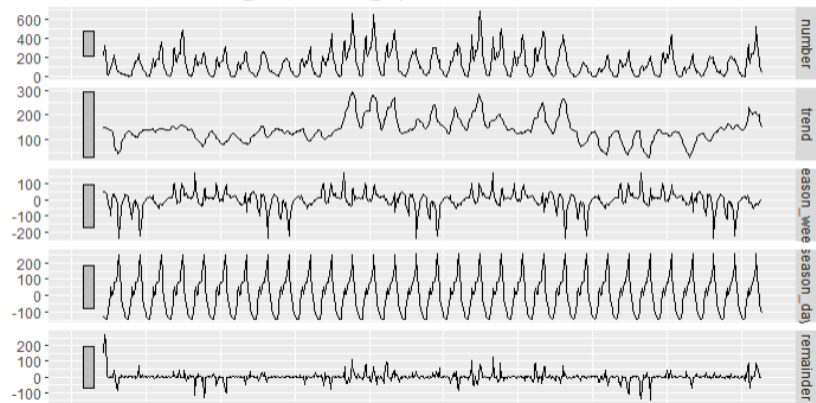


模型基本能够反映在不同时点上，其是处于用车的高峰期还是低估期。然而，模型依旧具有较大的误差，并不能够精确预测一个具体时点上的用车数量。出现这一情况的主要原因在于，人们在周中和周末的用车习惯不同导致数据样本具有两个周期性，从而导致 ARIMA 模型无法准确预测。通过 STL decomposition 可以很好地看出这一点：



STL decomposition

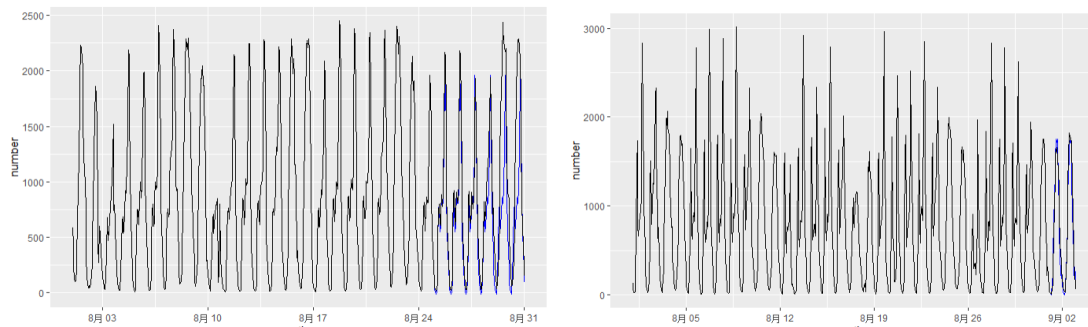
number = trend + season_week + season_day + remainder



为能够更加准确地进行预测，我们将数据分为周中和周末，分别进行分析。

2.2.3 周中与周末数据的预测

选用同样的模型对周中与周末数据进行预测，结果如下：



上图为周中数据的预测结果，下图为周末数据的预测结果。分样本后，预测精度有了显著的提升，能够较好地反映数据的真实情况。结果证明，周末用车时段较为分散，周中则主要集中于上下班高峰。

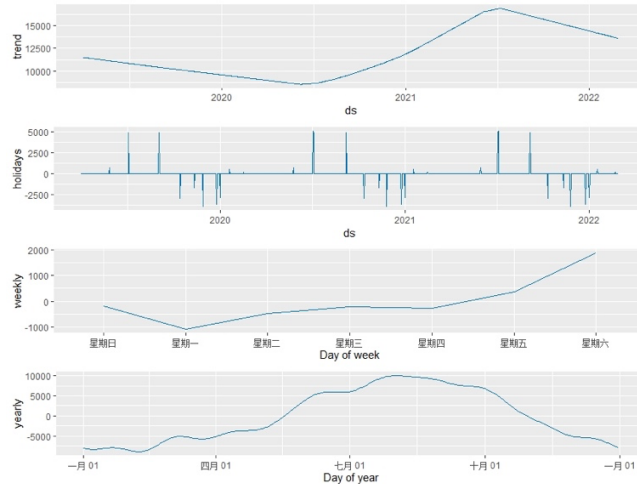
三、Prophet Model

考虑到 ARIMA 的经典模型无法对数据中多重的周期性进行学习，我们进一步尝试对该数据集应用 Prophet 模型。Prophet 模型作为可分解的时间序列模型，主要由趋势项，季节项和假期项组成，且该模型可以识别出数据中多重的周期性，并具有拟合速度快、无需进行缺失处理等特点。

3.1 天级数据拟合与预测

在天级数据层面，我们主要利用数据集中前 35 个月作为训练集，后 1 个月作为测试集。我们在模型中加入了年和周两种季节性，且加入了法定节假日作为假期项。

具体 component 的分解如下：

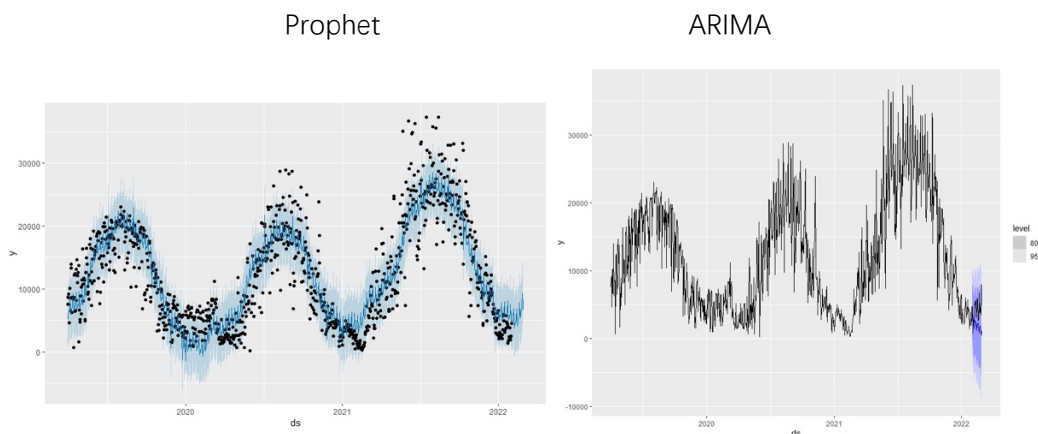


从图中我们可以发现, Prophet 模型能够较好的识别天级数据中的周与年的双重周期性, 比如一年中不同季节对用车的影响等等。其次, 加入的公众假期的效果也十分显著。但是考虑到夏季用户的数量明显更多, 我们猜测存在周的周期性被部分吸收的可能。

在测试集的表现方面, 我们主要选取了最优化搜寻下的 ARIMA 作为比较。从下方的表格可以看出, 在天级数据的较长期预测中, 各类误差度量下 Prophet 的表现显著更好。我们认为主要是由于 ARIMA 较为依赖预测窗口前的数据且无法识别天级别的多重周期性。

Model	MSE	RMSE	MAE	MAPE
Prophet	2983625	1727	1680	27.6
ARIMA (1,1,1) (1,0,2) [7]	8561476	2926	2482	65.7

此外, 从下面的图中可以更明显地发现, ARIMA 模型较为依赖先前的数据, 给出了持续向下的趋势预测; prophet 模型较好的捕捉到了季节性的变化, 预测表现与真实数据趋势较为一致。

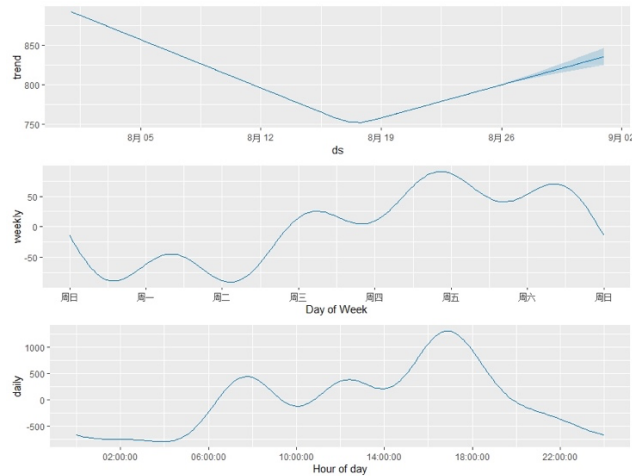


3.2 小时级数据拟合与预测

我们同样尝试在小时级别上应用 Prophet 模型进行学习, 主要选取了 8 月的前一段时间作为训练集, 而最后一周作为测试集。

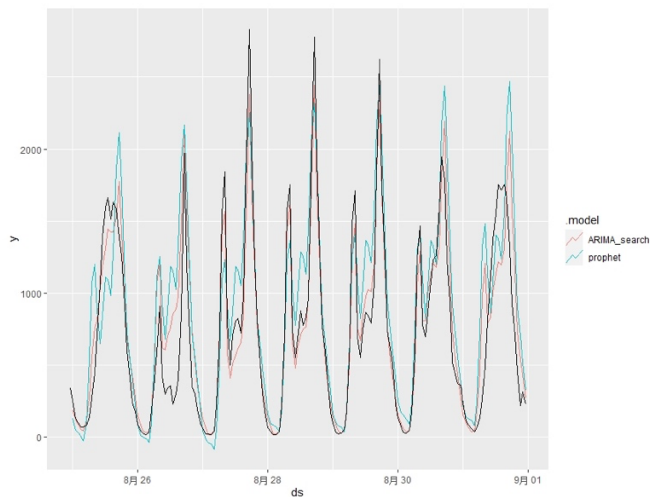


从 component 的分解图中我们可以看出，prophet 模型可以较好的识别出早高峰和晚高峰的特征，如早上 8 点和下午 5 点左右的波峰。



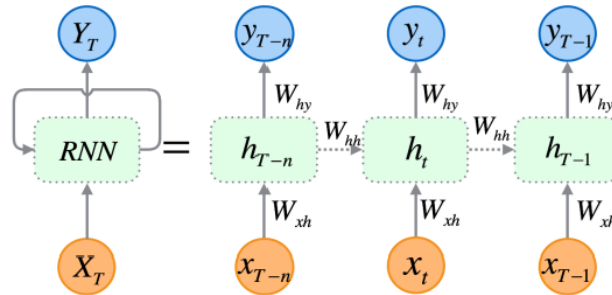
在测试集上，我们同样选取了 ARIMA 进行比较，从下方的表格中可以看出，在小时级的较短期预测中，ARIMA 的表现要更好；我们猜测小时级别的变化可能更依赖于预测窗口前的数据，例如前几个小时的天气情况对随后的用车量的影响，而 ARIMA 的特性可以较好的捕捉这一特点。

Model	MSE	RMSE	MAE	MAPE
Prophet	140625	375	276	82.1
ARIMA (0,0,2) (2,1,2) [24]	55225	235	161	38.9

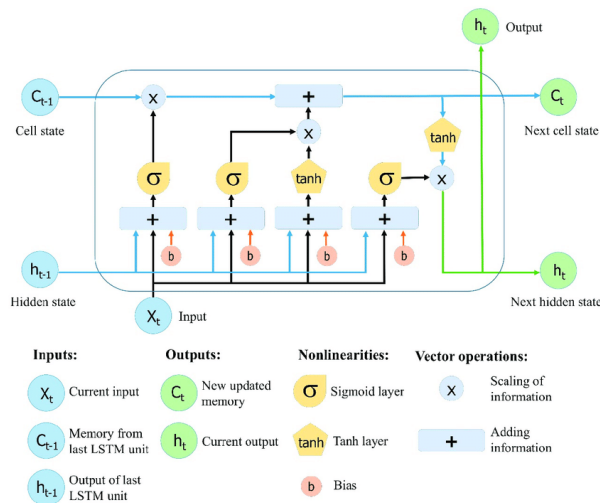


四、LSTM-RNN Model

Rnn 和全连接神经网络不同之处在于它的隐藏层之间是具有联系的，上一个 X_{t-1} 对应的 h_{t-1} 会通过 $W_{h_t h_{t-1}}$ 影响下一个 h_t 的输出，也是因为这一层影响他可以用来进行时间序列数据的预测，但同时它还有一个问题，就是所有的 RNN 隐藏层结构使用的都是同一组 (W_{hh}, W_{xh}, b) ，这也导致普通 RNN 模型不具备长期记忆性，因为误差在反向传播到距离很长时间维度的点上时很容易发生梯度消失或者梯度爆炸。



但是 LSTM 通过引入门机制和长期记忆细胞就比较好的解决了这类问题。第一个遗忘门用来控制输入的 x 和上一层隐藏层的输出 h 被遗忘的程度；输入门控制 x 和当前计算状态更新到记忆单元的程度大小，这两个门的输出对记忆细胞状态进行更新，最后再由当前细胞通过变化后和当前输入情况加和得到最终的输出，并作为下一个时点的输入结点。

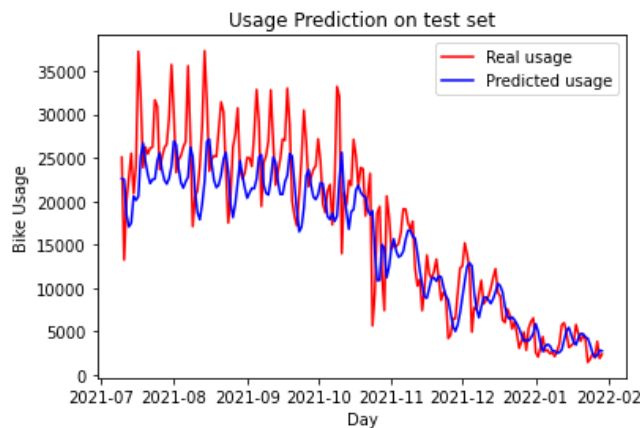


4.1 One-step recursive forecast

4.1.1 天级数据的拟合和预测

对于天级数据，我们留了 3 年中的最后一个月用于单步迭代预测，剩下的数据集前 80% 用于 LSTM 的训练，后 20% 用于测试。

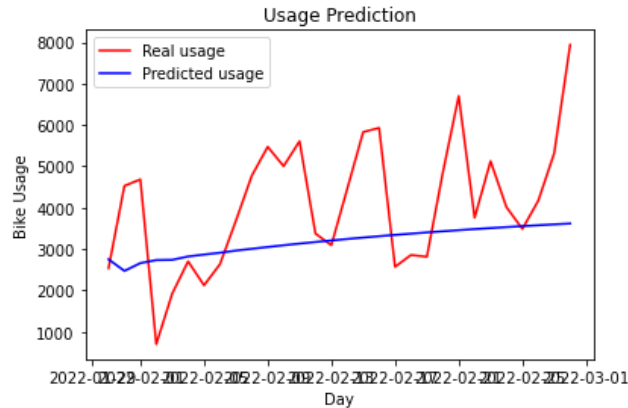
通过不断地调整输入天数、隐藏层个数、节点数，并通过网格搜索选择超参数之后，我们最终选择用前 3 天的数据对未来 1 天进行预测，在测试集上的拟合效果如下图所示：



可以发现整体的拟合效果不错，测试集上的 MAPE 达到了 25.02，但是前期的天际较大波动并没有能很好地捕捉住。但是测试集是依赖真实数据作为输入的，接着我们继续对最后



一个月的数据进行单步迭代预测，预测的效果如下图所示：

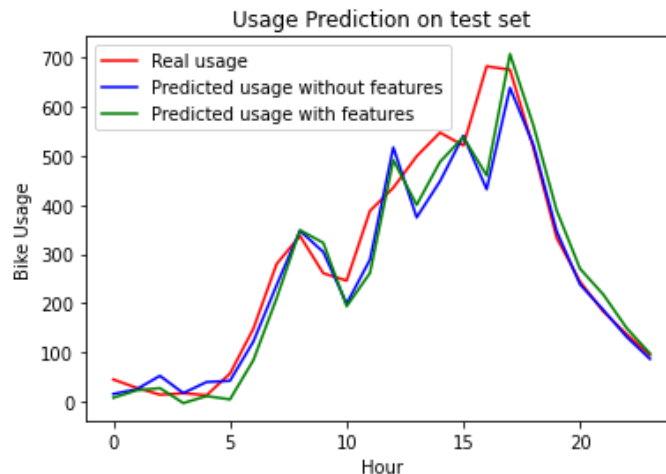


预测的结果非常差，仅仅展示了长期的趋势性，日间的波动情况完全被忽略，可能的原因应该是在单步迭代预测的过程中误差被不断放大。

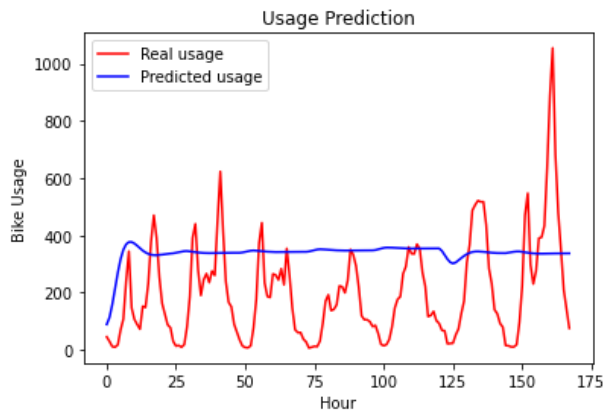
4.1.2 小时级数据的拟合和预测

小时级数据部分，我们总共选取了最后 5000 个小时，并留了最后的 168 (7*24) 小时作为单步迭代预测的比较，剩下的数据也同样是前 80%用于训练，后 20%用于预测。

这个部分我们尝试了两种输入的形式，第一种是仅输入一维的时间序列数据并输出；第二种是输入二维的数据（时间序列数据和该小时是星期几的特征变量）并输出一维的时间序列数据预测。我们想要考察当加入星期几作为特征变量告诉神经网络后能否提高整体的预测精度。同样地，通过不断地调整输入长度、隐藏层个数、节点数，并通过网格搜索选择超参数之后，我们最终选择用前 12 小时的数据对未来 1 小时进行预测，在测试集上的拟合效果如下图所示：



总体来看，两种输入形式的拟合效果都还不错，在中期的波动相较真实数据多一些；对比来看，一维输入的 MAPE 为 69.82，而二维输入的 MAPE 为 45.47，可见加入作为星期几这个特征确实能够在一定程度上加强神经网络的学习能力，达到更好的拟合效果。接着我们基于二维输入继续对最后 168 小时的数据进行单步迭代预测，预测的效果如下图所示：



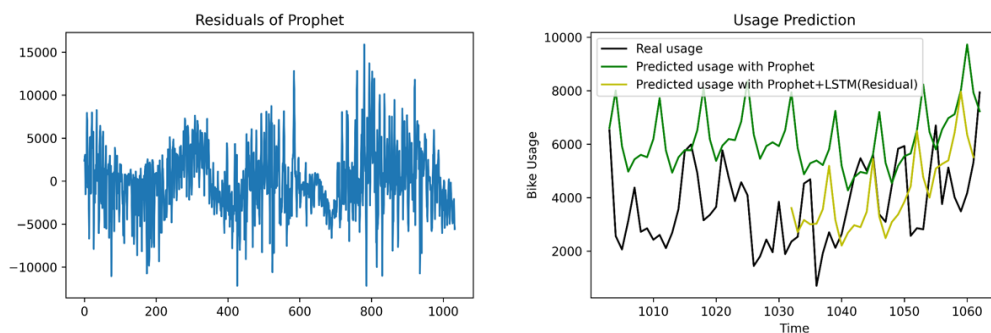
预测效果仍然很差，仅在前 12 小时表现出了小时间的波动性，随着时间维度的拉长预测误差也被不断放大，最终转为了一条水平线的预测结果。为了改善预测效果和误差放大的问题，我们基于 LSTM 做了如下尝试。

4.2 集成残差学习（LSTM+Prophet）

Hansika Hewamalage (2020) 发现 RNN 在时间序列存在同质季节性时能够有较好的表现，否则需要进行去季节性的操作。而我们的数据集明显存在多重的季节性。天级数据有星期和年份的双重周期性，并且星期的周期性并不是很规律。我们认为 LSTM 在多步天级中表现较差的原因正是如此。

而 Prophet 模型关注于多周期性以及趋势性，忽略了自相关性。即便如此，Prophet 在天级的多步预测中表现依旧是以上模型中最出色的。我们希望发挥 LSTM 模型的优势，进一步提升 Prophet 模型的预测效果。

方法便是先使用 Prophet 模型在训练集（前 35 个月）进行拟合，与真实值的差距得到残差。将残差作为学习对象，用 LSTM 进行训练。预测时用两个模型分别进行多步预测，并将结果相加作为预测。



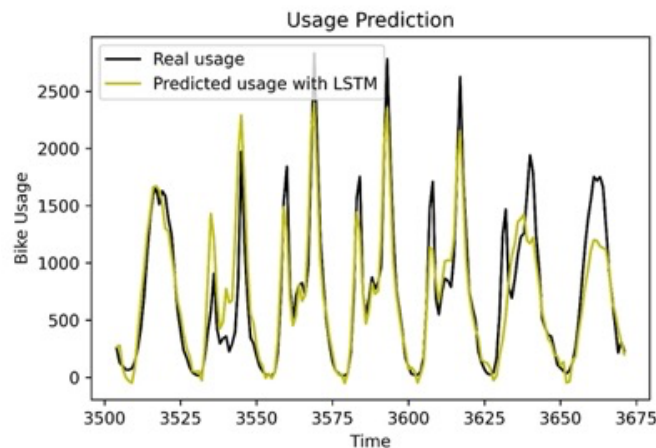
我们可以看到，结合了 LSTM 预测的残差，Prophet 模型的效果显著的变好了。MAPE 也从 88.972 下降为 56.058。通过以上结合，也说明了集成学习的效果。

4.3 Moving Window 多步预测（LSTM）

我们先前在小时多步预测中发现效果并不好。但我们认为小时数据与天级数据不同，存在着非常强的同质周期性（24 小时），理应是可以通过 RNN 进行预测。我们认为可能是输入的规格以及单步 recursive 的问题。单步 recursive 会使得预测误差不断累积，我们要预测 7*24 步，显然效果会受到很大的影响。而如果直接预测未来 7*24 小时的模型，显然对于模

型的拟合要求很高，很有可能残差不收敛，无法学习成功。我们认为以 24 小时为一个周期进行每小时的预测是十分合理的。因为从数据来看，24 小时中的每个时刻都有鲜明的特征，去预测未来 24 小时每个时刻的结果，就能很好的去学习到一天的波动规律。另外以 24 小时进行滚动预测能保证后续的数据能大致符合之前真实的规律，作为后续的训练输入能保留周期性。

输入则选择了过去 7×24 小时，因为如果需要预测到未来的某天是否是周末，如果提供足够的信息（过去的 7 天数据）那么是能够反映周末与否的。并且预测时，上一周的同天数据也能提供很多的信息。



对于未来 7×24 步的预测，模型只需要跑 7 次循环。将预测的 24 小时作为新的输入的末尾，并且去除上一次输入开始的 24 小时。可以看到，预测的效果非常的好。效果不只是体现在拟合程度高，并且还有有效的预测出周末的波峰在下午，周中的双波峰在早晚高峰。说明过去 7×24 小时的输入的信息是有效的。