Forecast Auto-Adjustment

版本	修订信息	作者	日期
v1.0	创建文档	王晶	20201105

This is a research and implement of auto-adjustment for demand forecast in rolling predict.

在滚动预测中，根据前面滚动的情况或者在调度过程中根据新加入的真实值，去调整我们模型的预测值，以实现更高的精确度。可用在高峰预测等场景。主要是根据最近的误差表现来修正模型的预测值输出，作为新的输出。

1 Error as Feature

Original Model: $f : \hat{y}_t = f(X_t)$

New Model: $f : \hat{y}_t = f(X_t, e_{t-h}), e_{t-h} = y_{t-h} - \hat{y}_{t-h}$

The simplest method is to add error in previous rolling as feature in current rolling. The inital error could be set 1.

最直接的方式是把上一轮滚动的预测误差e_{t-h}作为下一轮的特征值加入。

2 Error Postprocess

Assume the fitted model is going to perform as previous. For example, if the model underestimates in $t-l$ predicting t, then it will do the same thing in t predicting $t+l$。

误差后处理方式。基于的假设是模型在该轮滚动的表现会延续上一轮的表现。比如上一轮滚动低估，该轮还是会低估。上一轮滚动高估，该轮还是会高估。

Let U be the event of model underestimation , O be the event of model overestimation.

让U是模型低估事件，O为模型高估事件。

该方法假设：$P(U_t|U_{t-l}) = 1$和$P(O_t|O_{t-l}) = 1$

Proof:

我们来证明这一假设是有一定合理性的。假如我们选用线性模型开始$\hat{y} = X\theta$。

$$\hat{y}_{t-l} = X_{t-l}\theta_{t-l}$$
$$\theta_{t-l} + \Delta\theta = \theta_t$$
$$\Delta\theta = -\alpha\frac{dloss_{t-l}}{dX_{t-l}}$$
$$\hat{y}_t = X_t\theta_t = X_t(\theta_{t-l} + \Delta\theta)$$

假设有个完美的模型$y = X\theta_p$，$loss = |y - \hat{y}|$。如果$\hat{y}_{t-l} \leq y_{t-l}$, 那么$\theta_{t-l} \leq \theta_p$ （如果$X_{t-l} > 0$，表示X_{t-l}是正定矩阵），前面低估，我们要证明后面也很可能低估。

$$\Delta\theta = -\alpha\frac{d(y_{t-l} - X_{t-l}\theta_{t-l})}{dX_{t-l}} = \alpha\theta_{t-l}$$
$$\hat{y}_t = X_t(1+\alpha)\theta_{t-l} \le X_t(1+\alpha)\theta_p = (1+\alpha)y_t \quad \text{如果}\, X_t > 0$$

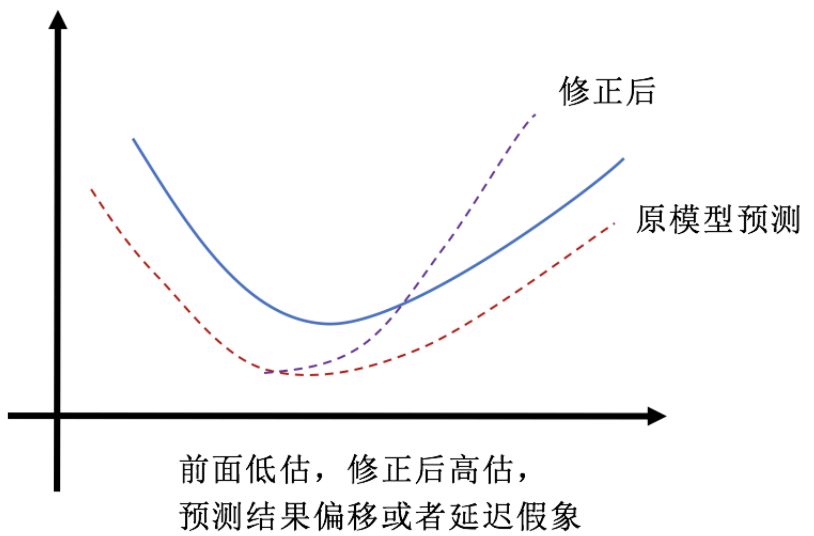
如果$\hat{y_{t-l}} > y_{t-l}$，即$\theta_{t-l} > \theta_p$。

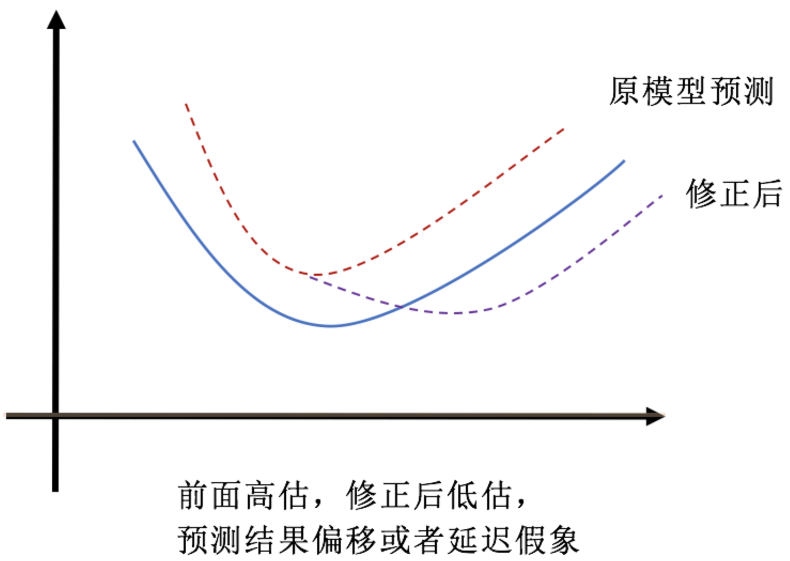
$$\Delta\theta = -\alpha\theta_{t-l}$$
$$\hat{y}_t = X_t(1-\alpha)\theta_{t-l} > X_t(1-\alpha)\theta_p = (1-\alpha)y_t$$

由于α比较小，我们可以近似不等式成立。$P(\hat{y}_t \le y_t|\hat{y}_{t-l} \le y_{t-l}) \approx 1$, $P(\hat{y}_t > y_t|\hat{y}_{t-l} > y_{t-l}) \approx 1$ 。α越小，也就是梯度更新越慢，假设越有可能成立。上述基于线性模型的情况下成立，或者当l相对较小的时候，我们可以认为y_{t-l}和y_t之间的接近线性的。但该假设还不能推广到更通用的情况。

当特征矩阵X_t是正定矩阵，且l较小的时候，假设大概率成立。

该方法应用：$e_{t-l} = y_{t-l} - \hat{y}_{t-l}$, $\tilde{y}_t = \hat{y}_t + e_{t-l}$，$\tilde{y}_t$为修正后的预测结果。在实际预测中会出现两种情况，造成看起来预测偏移延迟的情况。





总体准确率会比后面不低估也不高估更高，因为出现误差抵消。

为了可能减少这种情况，但不一定能够提升预测指标。我们引入趋势项进行规则调整。

- 如果训练集末尾趋势是增长的
    - 前面高估的部分沿用模型输出的原始结果，低估的部分进行误差修正
- 如果训练集末尾趋势是降低的
    - 前面低估的部分沿用模型输出的原始结果，高估的部分进行误差修正

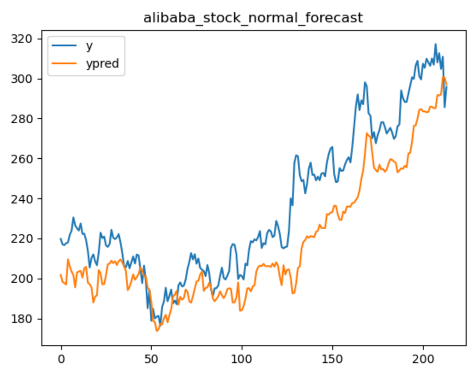
3 Practice

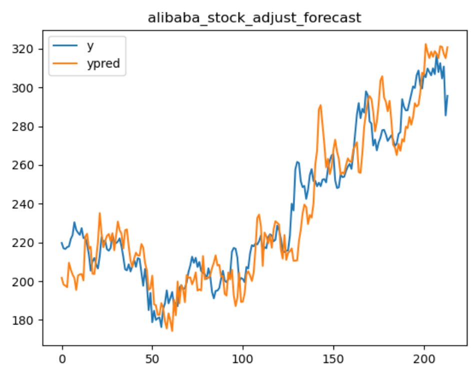
我们选用Amazon，Google，Alibaba和JD四家公司两年的股票收盘价作为测试数据验证，原始预测，误差修正和误差趋势修正的对比结果。对比指标使用$MAPE$和CV。模型采用线性模型，特征采用一系列事件序列特征。结果如下图

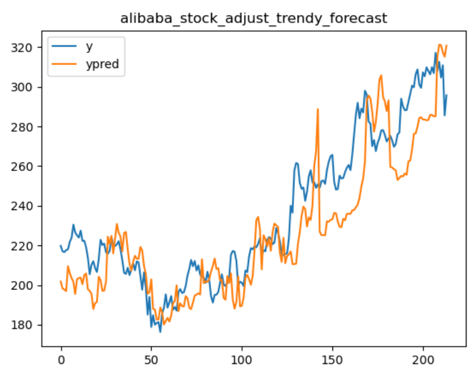
公司	原始预测平均MAPE	原始预测CV	误差修正平均MAPE	误差修正CV	误差趋势平均MAPE	误差趋势CV
Amazon	0.103	**0.608**	**0.065**	0.638	0.075	0.776
Google	0.076	**0.650**	**0.057**	0.766	0.066	0.768
Alibaba	0.082	0.450	**0.048**	**0.407**	0.069	0.436
JD	0.062	0.597	**0.045**	**0.437**	0.053	0.474

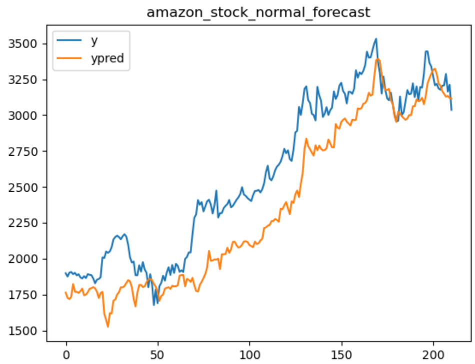
误差趋势的方式从效果表现来看没有误差修正好，但比原始预测要好一些。主要原因是如前所述，在计算平均MAPE和CV的时候，误差修正产生了更多的误差抵消。

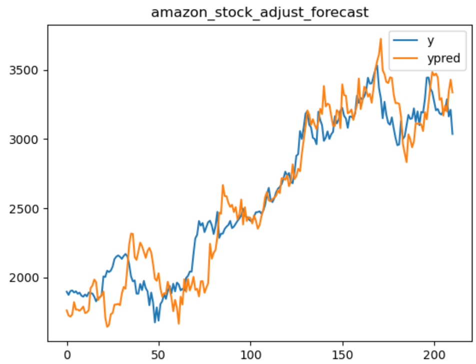
示意图：

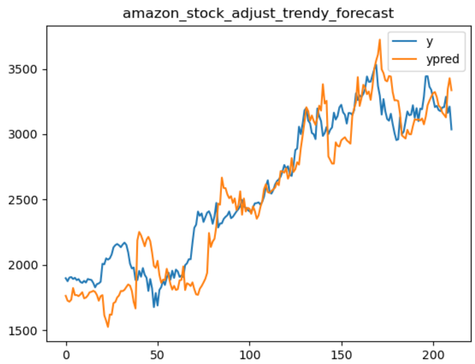


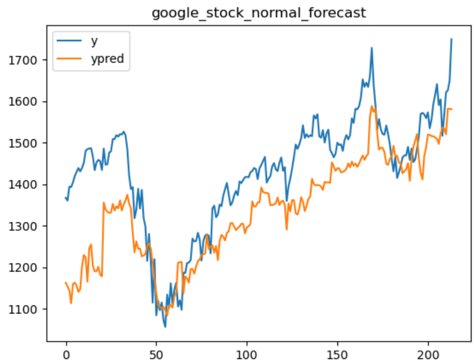


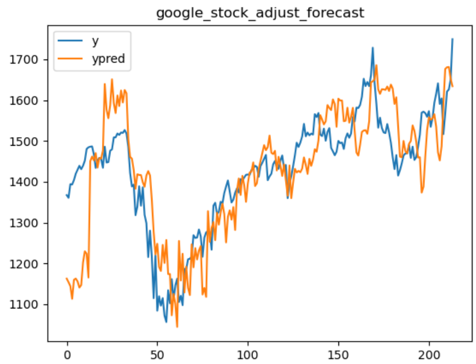


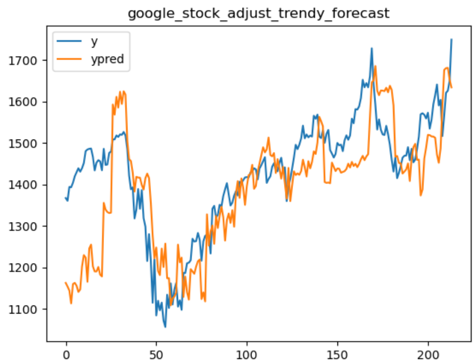

amazon_stock_normal_forecast
y
ypred
3500
3250
3000
2750
2500
2250
2000
1750
1500
0
50
100
150
200

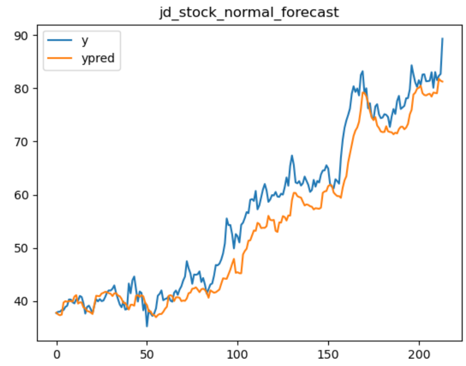

amazon_stock_adjust_forecast
y
ypred
3500
3000
2500
2000
0
50
100
150
200

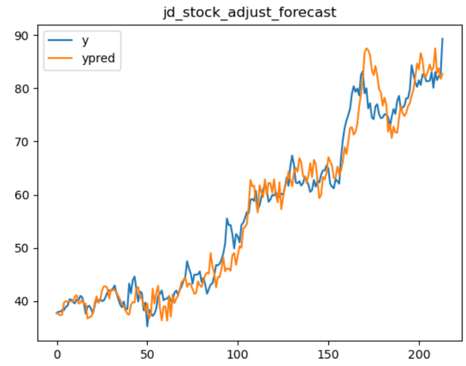

amazon_stock_adjust_trendy_forecast
y
ypred
3500
3000
2500
2000
1500
0
50
100
150
200

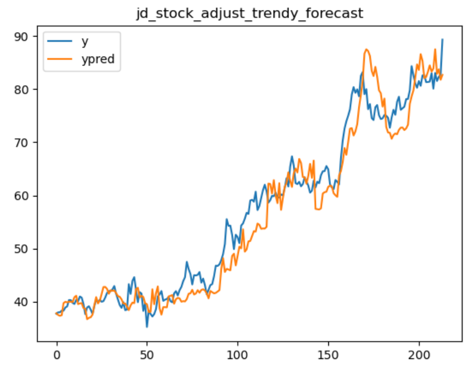

google_stock_normal_forecast
y
ypred
1700
1600
1500
1400
1300
1200
1100
0
50
100
150
200


google_stock_adjust_forecast
y
ypred
1700
1600
1500
1400
1300
1200
1100
0
50
100
150
200


google_stock_adjust_trendy_forecast
y
ypred
1700
1600
1500
1400
1300
1200
1100
0
50
100
150
200


jd_stock_normal_forecast
y
ypred
90
80
70
60
50
40
0
50
100
150
200


jd_stock_adjust_forecast
y
ypred
90
80
70
60
50
40
0
50
100
150
200


jd_stock_adjust_trendy_forecast
y
ypred
90
80
70
60
50
40
0
50
100
150
200