

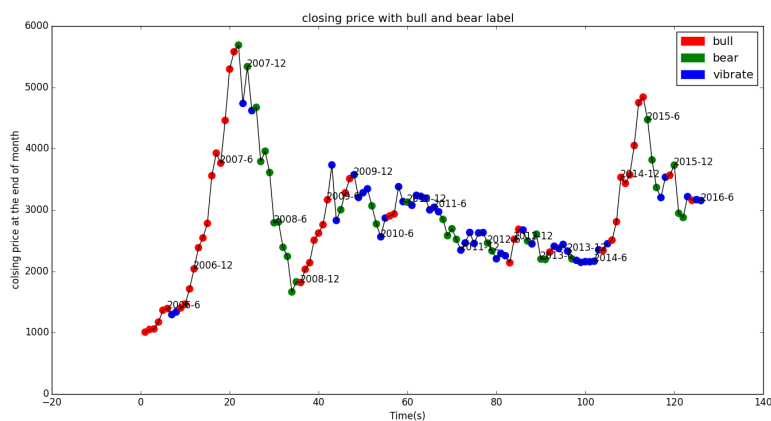
牛熊市试验报告 IV

邱明轩*

08, 19, 2016

1 Visualization

为了更好地做试验、改进模型，我把数据可视化了。可视化利用了月末的牛熊标签以及收盘价。牛熊标签的计算方法如下：利用某日两个月之前以及一个月之后的收盘价，计算月化收益率，如果收益率超过3%，标注为牛市；如果收益率低于-3%，标注为熊市，中间情况为震荡市。可视化结果如下：



2 Result

接下来我把试验结果可视化了，针对试验数据不足的情况（只有126个试验数据），我采用了留一交叉验证。具体的讲，就是每次用125个点训练模型，预测预留的点，循环126次。伪代码以及实验结果如下：

伪代码

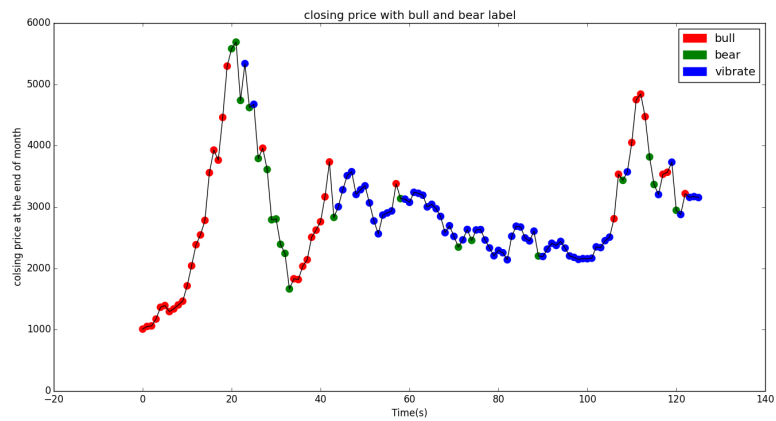
```
1 for i in ranges(0,126):
2     for j in ranges(0,126):
```

*email: mingxuandi@163.com

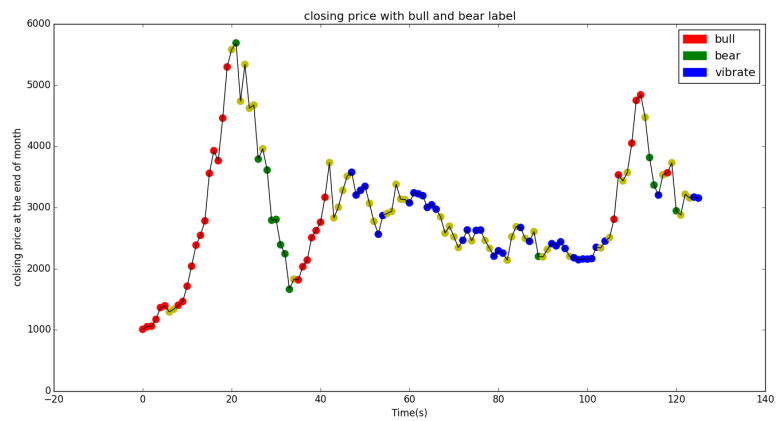
```

3         train data list.
4         if j==i:
5             get test data
6         else:
7             add data to train data list.
8     end
9     train model using train data
10    predict test data using trained model
11 end

```



利用留一验证方法得到的每个月的预测结果



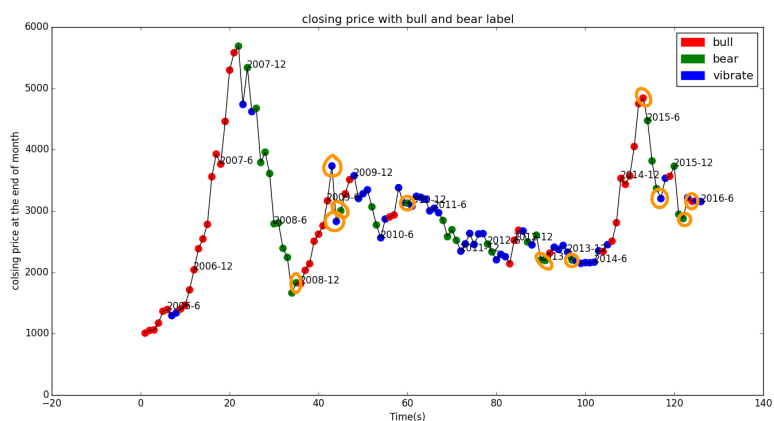
与数据标签不同的预测结果用黄色标注出来了

从图中可以看出，模型在大牛市大熊市中表现很好，在10年到14年震荡的情况中表现较差。模型可以抓住大的机会，规避大的风险，但是对于小机会，小风险，把握能力不强。感觉目前的市场行情比较符合10年的情况，模型可能会一直判断为震荡市，针对这种情况，比较好的方法是利用10年到14年的数据重新训练一个模型，利用老模型提示大机会大风险，利用新模型重新分析老模型判断为震荡的情况，提高模型整体的效果。

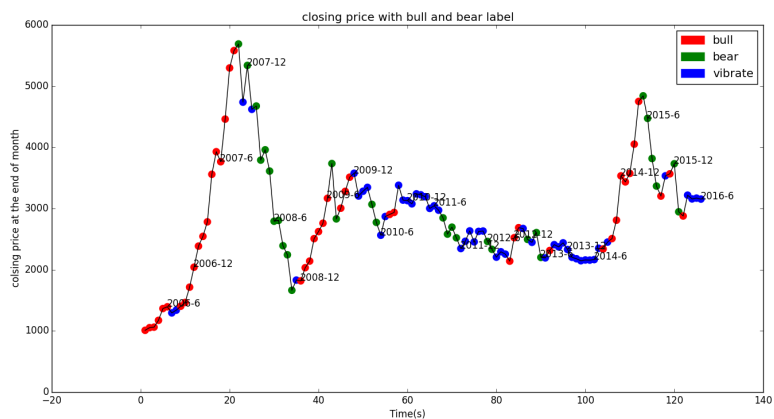
3 Improvement

提高模型的效果，一个比较简单的方法是在数据上做文章，包括增大数据量以及提高数据的纯度。现在我们尝试提高数据的纯度。之前机器利用3%方法自动标注的数据不一定准确，我们利用主观判断，手动更改数据的标签，使标注结果更加符合我们的要求。标注结果如下图，我们希望得到的模型的预测效果能与这些标注尽量相同。

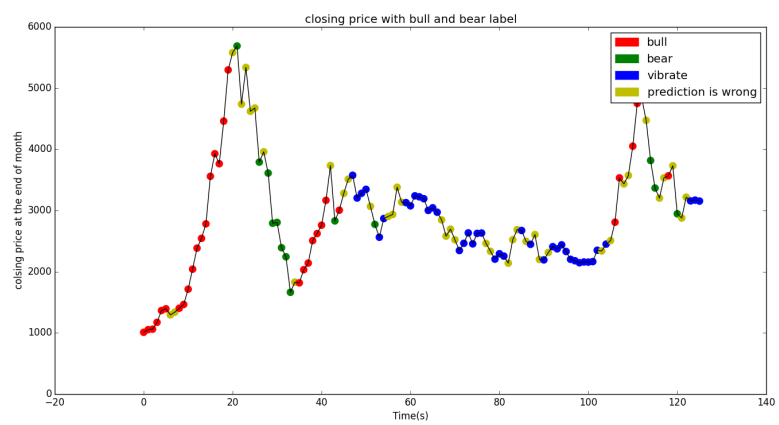
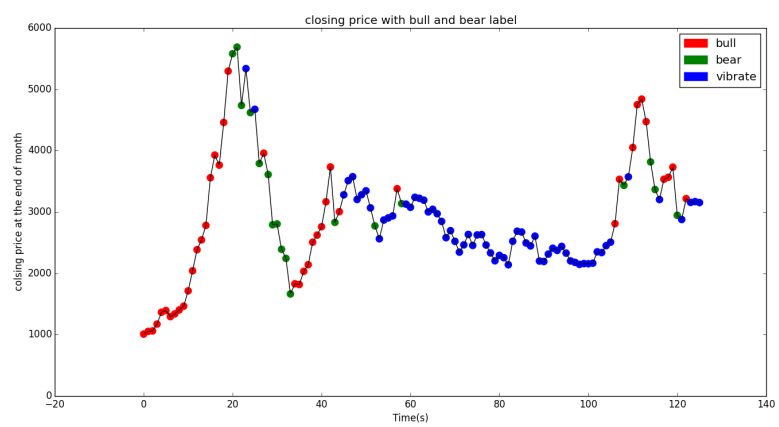
调整标签之后，预测的准确率有所提升，达到67.5



橙色为需要手动调整的数据



橙色为手动调整的数据



调整后标签的分布更加合理，标签分布与数据的规律更加一致，可以预测的更准确。