## Result

基于几个不同的对比方法，我们在TSLA、PAICC、MSFT三个数据集上进行了详细的实验分析，我们根据4.2中给出的多分类指标选取了Macro，Weighted，AUC三类指标，其中AUC包含了跌、平、涨每一类的分类性能描述，Macro和Weighted包含了对应的precision,recall,f1-score指标，详细的结果在表2，表3，表4给出，我们的方法对应表中的our GAN1和our GAN2，其中our GAN1是我们没有经过保序回归优化的模型，our GAN2是我们在加入保序回归思想优化后的模型。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 2 The experiment result on TSLA | | | | | | |
| Indicator | LSTM | GRU | CNN | Conv-  LSTM | our GAN1 | our GAN2 |
| Class0 AUC | 0.5094 | 0.5395 | 0.5037 | 0.4831 | 0.5482 | 0.5212 |
| Class1 AUC | 0.6201 | 0.6126 | 0.5955 | 0.5482 | 0.6372 | 0.7698 |
| Class2 AUC | 0.5391 | 0.5124 | 0.4956 | 0.5014 | 0.5364 | 0.5218 |
| Macro-precision | 0.3451 | 0.3529 | 0.3803 | 0.3294 | 0.3695 | 0.3816 |
| Macro-recall | 0.3473 | 0.3518 | 0.4058 | 0.3822 | 0.4529 | 0.4707 |
| Macro-f1-score | 0.3087 | 0.3051 | 0.3761 | 0.2922 | 0.3495 | 0.3858 |
| Weighted-precision | 0.5199 | 0.5324 | 0.5182 | 0.4862 | 0.5332 | 0.5378 |
| Weighted-recall | 0.4814 | 0.4839 | 0.4888 | 0.3722 | 0.4615 | 0.5012 |
| Weighted-f1-score | 0.4439 | 0.4363 | 0.4843 | 0.4159 | 0.4713 | 0.5133 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 3 The experiment result on PAICC | | | | | | |
| Indicator | LSTM | GRU | CNN | Conv-  LSTM | our GAN1 | our GAN2 |
| Class0 AUC | 0.5603 | 0.4890 | 0.5663 | 0.4389 | 0.5283 | 0.5759 |
| Class1 AUC | 0.6621 | 0.5867 | 0.6611 | 0.3216 | 0.6131 | 0.6647 |
| Class2 AUC | 0.5427 | 0.4950 | 0.5028 | 0.4904 | 0.5271 | 0.5228 |
| Macro-precision | 0.2940 | 0.3634 | 0.3148 | 0.2887 | 0.4041 | 0.3699 |
| Macro-recall | 0.3351 | 0.3362 | 0.3365 | 0.3251 | 0.3876 | 0.3600 |
| Macro-f1-score | 0.2229 | 0.3314 | 0.2257 | 0.2781 | 0.3914 | 0.3436 |
| Weighted-precision | 0.3850 | 0.4031 | 0.4120 | 0.3780 | 0.4462 | 0.4285 |
| Weighted-recall | 0.4442 | 0.4212 | 0.4462 | 0.4288 | 0.4558 | 0.4635 |
| Weighted-f1-score | 0.2951 | 0.4041 | 0.2987 | 0.3657 | 0.4488 | 0.4348 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 4 The experiment result on MSFT | | | | | | |
| Indicator | LSTM | GRU | CNN | Conv-  LSTM | our GAN1 | our GAN2 |
| Class0 AUC | 0.5129 | 0.5358 | 0.5221 | 0.5220 | 0.5228 | 0.517 |
| Class1 AUC | 0.5440 | 0.4930 | 0.5322 | 0.4856 | 0.5386 | 0.5703 |
| Class2 AUC | 0.5273 | 0.5632 | 0.5256 | 0.5310 | 0.5051 | 0.5724 |
| Macro-precision | 0.3506 | 0.3179 | 0.3575 | 0.3438 | 0.3654 | 0.3639 |
| Macro-recall | 0.3528 | 0.3450 | 0.3585 | 0.3425 | 0.3649 | 0.3669 |
| Macro-f1-score | 0.3134 | 0.3011 | 0.3519 | 0.3400 | 0.3410 | 0.3584 |
| Weighted-precision | 0.3670 | 0.3407 | 0.3690 | 0.3588 | 0.3804 | 0.3784 |
| Weighted-recall | 0.3664 | 0.4040 | 0.3597 | 0.3450 | 0.3423 | 0.3664 |
| Weighted-f1-score | 0.3299 | 0.3414 | 0.3575 | 0.3490 | 0.3363 | 0.3664 |

从表2，表3，表4中我们得知经过保序回归优化后的模型our GAN2在大部分指标上的表现都优于已有的深度学习方法，特别在PAICC数据集上除Class2 AUC指标我们的方法要优于其他方法，并且在数据集TSLA上，our GAN2在Class2 AUC上的指标达到了0.7698，相比于对比方法中最高的0.6126提升了0.1572，说明我们的方法分类精度得到了提升。同时在其他的分类指标中，Macro-precision指标提升最多的是在数据集MSFT上为0.0064，Macro-recall指标提升的最多的是在数据集PAICC上为0.0235，Macro-f1-score指标提升的最多的是在数据集PAICC上为0.0122，Weighted-precision指标提升最多的是在数据集PAICC上为0.0165，Weighted-recall指标提升最多的是在数据集PAICC上为0.0173，Weighted-f1-score指标提升最多的是在数据集PAICC上为0.0307，需要注意的是上述数据描述都是在每个指标上选取其他方法中表现最好的与我们的方法进行对比，在每个数据集上基于单独的指标我们的方法会有更大的提升。我们的生成对抗模型中的生成器是ConvLSTM，从表2，表3，表4中可以看到our GAN1，our GAN2在数据集TSLA上的9种指标上都有所提升，our GAN1在数据集PAICC上6个指标有所提升，在数据集MSFT上有两个指标有所提升，可见ConvLSTM作为生成器加入到生成对抗网络中，在分类性能上相比于端到端的ConvLSTM有所提升。并且我们加入保序回归思想优化后的our GAN2模型相比our GAN1模型在分类结果上有所改善，我们在表5中给出在实验数据集上的混淆矩阵结果。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 5. The results of confusion\_matrix in the datasets. | | | | | | | | |
|  | TSLA | | | | | | | |
|  | Our GAN1 | | | | Our GAN2 | | | |
| True Label |  | Up | O | Down |  | Up | O | Down |
| Up | 104 | 7 | 63(0.90) | Up | 93 | 10 | 71(0.87) |
| O | 3 | 2 | 0 | O | 3 | 2 | 0 |
| Down | 135(0.94) | 8 | 81 | Down | 102(0.87) | 15 | 107 |
|  | Predicted Label | | |  | Predicted Label | | |
|  | PAICC | | | | | | | |
|  | Our GAN1 | | | | Our GAN2 | | | |
| True Label |  | Up | O | Down |  | Up | O | Down |
| Up | 148 | 4 | 78(0.95) | Up | 109 | 9 | 112(0.92) |
| O | 22 | 2 | 42 | O | 34 | 11 | 21 |
| Down | 128(0.96) | 5 | 91 | Down | 86(0.80) | 21 | 117 |
|  | Predicted Label | | |  | Predicted Label | | |
|  | MSFT | | | | | | | |
|  | Our GAN1 | | | | Our GAN2 | | | |
| True Label |  | Up | O | Down |  | Up | O | Down |
| Up | 88 | 116 | 42(0.26) | Up | 63 | 99 | 84(0.45) |
| O | 54 | 96 | 39 | O | 43 | 81 | 65 |
| Down | 113(0.47) | 126 | 71 | Down | 79(0.43) | 102 | 129 |
|  | Predicted Label | | |  | Predicted Label | | |

从表5中的混淆矩阵结果我们可以得知，我们最不希望出现的情况是跌预测为涨、跌预测为涨这两种情况，在数据集TSLA和PAICC上，我们加入保序回归思想后的模型our GAN2在上述两种情况下都比our GAN1有所提升，在测试集数据分布较均匀的MSFT数据集上our GAN2在避免将涨预测为跌这一情况比our GAN1表现稍差，我们猜测是由于该数据集的数据分布更适合our GAN1模型，因此在该情况下的表现由于our GAN2。但是从整体上来看，进过保序回归优化后的模型our GAN2在上述两种情况下表现要优于our GAN1，并且在整体的性能上有所提升，尤其是在Macro-f1-score指标上都有所改进。

## Conclusion

在金融时间序列价格的运动走势分类研究中，我们提出的改进的生成对抗网络模型相比已有的深度学习方法有了显著的改进，ConvLSTM模型在生成器中采用具有优秀的时间序列处理能力的卷积神经网络模型，实验结果表明它相比单独的CNN和CNN-LSTM在端到端分类中更加优秀，加入了生成对抗网络，使得CNN-LSTM模型的分类性能得到提升。保序回归思想的加入，使得我们的模型在分类结果的表现更具有现实意义，尤其是在将跌预测为涨，涨预测为跌这两种不利于现实中实际交易的情况下，our GAN2相比于our GAN1在上述情况下做了进一步的优化，使得我们的模型不仅在整体的分类性能上有所提升，而且给实际的交易提供了指导意义。