实验一：少标记样本下的二分类效果。数据集大小：417383，训练集：375644，测试集：41739，训练过程：epochs=50, batch\_size=128。

|  |  |  |
| --- | --- | --- |
| 100%标记的训练集：375644  Accuracy:100.00%  Recall:100.00%  Precision:100.00% | 10%标记的训练集：37564  Accuracy: 99.99%  Recall: 99.98%  Precision: 100% | 5%标记的训练集：18782  Accuracy:99.99%  Recall:99.98%  Precision:100.00% |
| 1%标记的训练集：3756  Accuracy:99.98%  Recall:99.95%  Precision:100.00% | 0.1%标记的训练集：375  Accuracy:89.03%  Recall:85.96%  Precision:88.27% |  |

由实验结果可以看出当标记样本的比例下降到0.1%时，HTFS的预测效果急剧下降，难以对未知样本进行准确判断。主要是因为模型可参考的训练集样本太少。但比CNN效果好，主要是因为HTFS除了使用流的原始特征作为输入外，还基于流的数据包序列提取了时间特征，并结合流的统计特征进行判断，因此具有更好的分类效果。

实验二：100%标记的恶意样本。从十类样本中选取一类作为测试类，不参与训练，总共重复十次。模型参数同1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 未知恶意家族 | 比例 | 准确率 | 精确率 | 召回率 |
| Cridex | 3.93% | 100% | 100% | 100% |
| Geodo | 9.81% | 100% | 100% | 100% |
| Htbot | 1.53% | 100% | 100% | 100% |
| Miuref | 3.23% | 100% | 100% | 100% |
| Neris | 8.10% | 100% | 100% | 100% |
| Nsis | 1.45% | 98.81% | 98.81% | 100% |
| Shifu | 2.31% | 100% | 100% | 100% |
| Tinba | 2.04% | 100% | 100% | 100% |
| Virut | 7.93% | 100% | 100% | 100% |
| Zeus | 2.63% | 100% | 100% | 100% |

由实验结果可以看出，HTFS模型对Nsis家族的识别能力最差，对该类别样本的泛化能力最弱，可能的原因是因为Nsis与其他恶意家族样本的特征存在较大不同，HTFS很难通过其他恶意家族的样本特征识别Nsis。或者是HTFS难以提取出Nsis和其他恶意家族区分度较大的特征。 比CNN的效果好，但是对于Nsis的识别效果没有VMT和MT的效果好。

实验三：少标记样本下的多分类效果，恶意样本：179252，标记样本数为10%：16132，训练集：17926。当epochs=50时，模型不容易收敛，因此设置epochs=200。

可以看出Neris的预测效果最差，主要是因为HTFS在模型预测过程中会把Neris误分类为Virut。

二分类和多分类实验同为10%标记样本，结果差异明显的原因：

二分类整体准确率为99.98%，多分类整体准确率为95.80%，训练集和测试集设置相同，两个实验结果差异明显的主要原因时，恶意流量与正常流量的区分比较明显，而恶意流量内部不同家族样本区分不明显。

少标记样本下HTFS的多分类效果比CNN稍差，二分类效果比CNN好很多，这反映了HTFS更擅长二分类任务。