基于数据集《vertical medical data》的 FATE 实验报告

蔡剑平

本实验采用了从 mimic III 中提取的数据子集，名为《vertical medical data》。数据集中包含了 3000 样本以及两方面数据内容，分别是“住院数据”和“门诊数据”。其中住院数据带有标签信息，表示是否存活，其中 1 表示患者样本存活，0 则表示患者不存活。

本实验采用纵向联邦学习方法训练两个参与方的模型。参与双方一边持有“住院数据”，一边持有“门诊数据”。其中，“住院数据”包含了 110 个属性加上标签，“门诊数据”包含了 133 个属性。

本报告实验了若干个纵向联邦学习的算法，并对各个算法在本数据集上的表现进行了评估。

1. 逻辑回归

本实验采用了 FATE 中的 HeteroLR 模块进行，实验代码见 py 文件“lr_vmd.py”，实验编号为 202204271056128136050。实验设置的迭代次数为 100 次，训练时长为 36 分 12 秒。训练过程的 Loss 变化如下图所示：

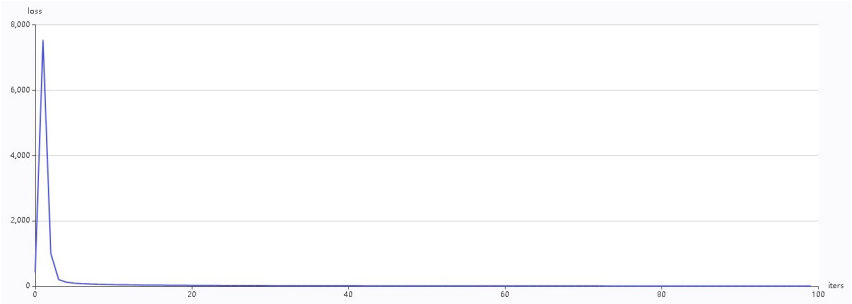

图 1.1 Loss 变化图

Loss 采用“L2”惩罚函数计算。从图 1.1 来看，HeteroLR 在第 1 次迭代后 Loss 急剧升高，但在后续的迭代过程中，Loss 稳步收敛，总体上来讲训练过程是成功的，实验的权值结果见“实验权重”表 1。训练结果如图 1.2-1.3 所示。

图 1.2 整体评价

图 1.3 ROC 图

从训练结果来看，整体评价的结果一般，未能达到理想的准确性。其中 AUC 值为 0.67，精度为 0.53，召回率为 0.65。在预测阈值方面，在阈值 0.5 下的预测准确率为 63.3%；在最佳阈值 0.47 下的预测准确率为 63.4%。

2. Boost 模型

2.1. 常规学习

本实验采用了 FATE 中的 HeteroSecureBoost 模块进行，实验代码见 py 文件“sbt_vmd.py”，实验编号为 202204271143397547850。实验设置为 5 棵树，划分深度最大为 10。训练时长为 4 分 45 秒，训练过程的 Loss 变化如下图所示：

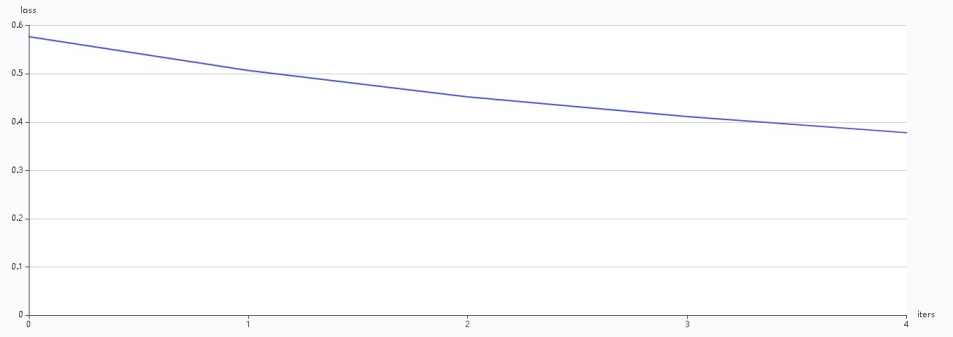

图 2.1 Loss 变化图

Loss 采用交叉熵函数计算。从图 1.1 来看，HeteroSecureBoost 在迭代过程中，Loss 呈现稳步收敛态势，总体上来讲训练过程是成功。训练结果如图 2.2-2.3 所示。

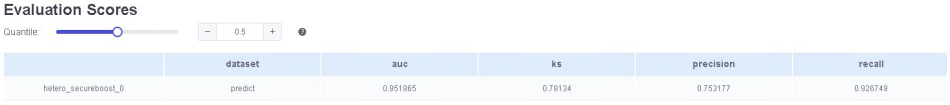

图 2.2 整体评价

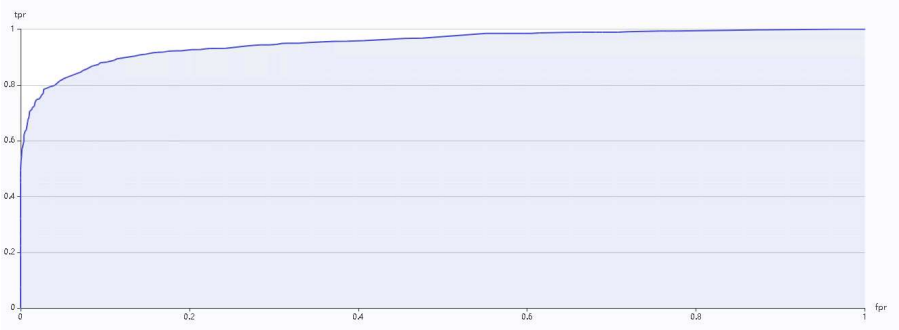

图 2.3 ROC 图

从上述实验结果来看，该实验获得了优秀的实验结果，AUC 值达到了 0.95，精度为 0.75，召回率为 0.92，并且 ROC 曲线也趋于饱满。在预测阈值方面，在阈值 0.5 下的预测准确率为 80.9%；在最佳阈值 0.47 下的预测准确率为 89.5%。

除了常规实验结果，本实验还获取了各个属性的特征重要性权值，如下图所示：

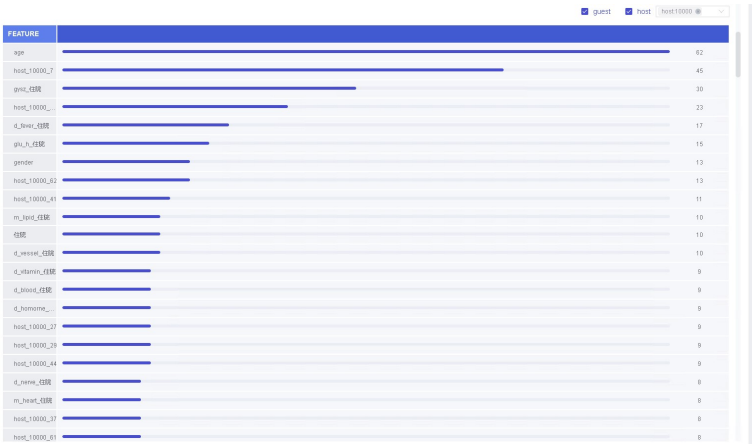

图 2.4 特征重要性

从 2.4 图来看，年龄是影响存活率的一个重要因素，其次是 Host 的属性 7，gysz_住院，host_10000_124…。

3. 神经网络模型

本实验采用了 FATE 中的 HeteroNN 模块进行，实验代码见 py 文件“nn_vmd.py”，实验编号为 20220427233743195196O，训练用时 53 分 21 秒。该实验采用了如下神经网络模型训练。

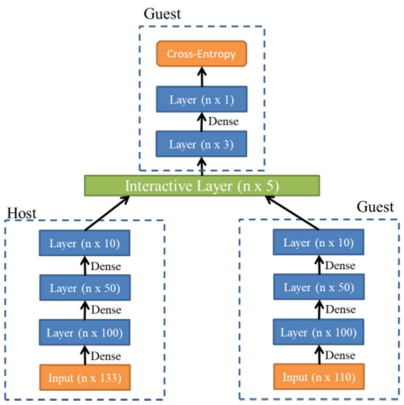

图 3.1 神经网络结构图

实验的 Loss 变化图如下图所示，可以看出，训练过程在前期并不是太稳定，存在波动。在大约 10 步左右开始收敛。由于 Loss 和 Boost 一样采用交叉熵函数计算，所以收敛结果可以看出并不如 Boost。

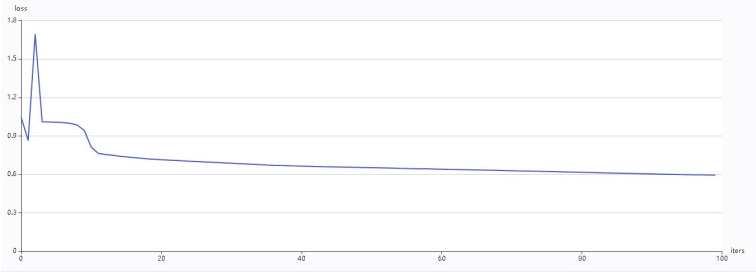

图 3.2 Loss 变化图

训练结果如图 3.3-3.4 所示。

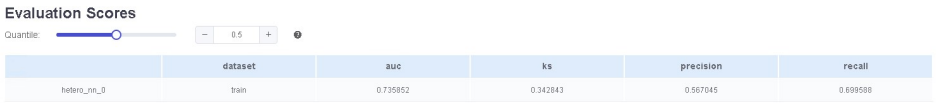

图 3.3 整体评价

图 3.4 ROC 图

从上述实验结果来看，该实验获得了相对于 LR 较好的实验结果，AUC 值达到了 0.73.5，精度为 0.57，召回率为 0.70。在预测阈值方面，在阈值 0.5 下的预测准确率为 67.7%；在最佳阈值 0.385 下的预测准确率为 68%。可以看出，最优阈值存在一定偏差。

4. KMeans 模型

本实验采用了 FATE 中的 HeteroKmeans 模块进行，实验代码见 py 文件“kmeans_vmd.py”，实验编号为 202204281405086066270。该实验将样本分为了 5 个类别，训练时长为 4 分 45 秒。实验的结果如下图所示。

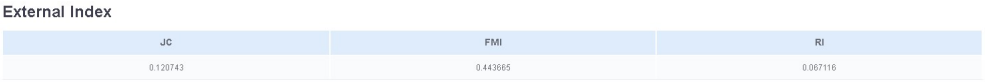

图 4.1 外部指标

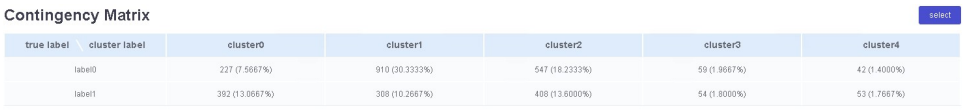

图 4.2 权变矩阵

从实验结果来看，三个外部指标{JC,FMI,RI}均比较低，没有达到 0.5。这初步说明联邦 KMeans 并不能很好地解决上述该分类问题。同时，通过权变矩阵可以看出，并没有哪个分类能够单纯地包含某个标签。

总结

从上述实验结果来看，联邦 Boost 算法能够取得比较好且稳定的实验结果。联邦逻辑回归的实验效果并不理想，说明所采用的数据集是非线性的。通过多次实验来看，神经网络模型的实验结果并不是太稳定，并且整体实验效果不如 Boost 模型。一方面，说明神经网络模型需要改进。另一方面，本文认为 HeteroNN 存在不稳定的因素。通过模型研究表明，Interactive Layer 作为该模型的关键，存在训练的梯度更新与其他子模型不同步的问题。因为它只支持最基础的 SGD 梯度更新，但 SGD 在实际应用中表现并不好。