1. 基本理论和方法（Basic Theory and Method）

GBDT是 Boosting 算法的一种，在每一轮预测中根据残差进行预测，最后将所有预测值相加得出结果。模型可以表示为决策树的加法模型

5．方法（Method）

作为上述方法的改进版本，LightGBM是一个更快速，性能更高的树集成模型。它将每个样本的一阶导数看做权重，梯度大说明距离最优解较远。在每次迭代的过程中，使用goss算法保留这些梯度大的样本点，而对于梯度小的样本点按比例进行随机采样，并且给小样本一个补偿系数，保证数据分布不发生改变。不断重复这一过程，使用采样得到的样本形成新的弱学习器，直到收敛或者达到规定迭代次数。

lightgbm有效避免了特征数多时数据切分不平衡的问题，它使用EFB算法将数据中的互斥特征绑定成为一个特征，从而减小特征维度，使模型的效率进一步提高。

它的Leaf-wise决策树生长策略可以更好地降低误差，但可能导致过拟合，所以在训练数据是要调整参数num\_leaves和max\_depth来得到最佳性能。

为了提升准确度，我们对预处理阶段得到的不同起点终点的线路分别训练模型。在具体实现选取learning\_rate=0.1，'num\_leaves': 31进行训练



