

手机游戏评分分类分析

摘要：本文研究了通过 15 项指标对安卓手机游戏评分进行分类的问题，构建了手机游戏分类评估模型，实现了高维度大样本条件下的综合定量分析。采用 Standard Scaler 正则化去除手机游戏数据大量有偏分布和极端值，只保留数据的排序关系。用 5 折交叉验证方法优化参数，网格搜索进行超参数整定，得出最优参数。将逻辑回归模型、SVM 和决策树等 7 种分类器结果比较，通过混淆矩阵直观展现分类效果，得 SVM 模型有最高的准确率和精确度，逻辑回归有最高的召回率。进一步，比较 Adaboost、投票分类器等 4 种集成算法，得出 Adaboost 准确率最高为 84.16%，故选取为最终模型。最后根据分类结果，为手机游戏评分的提高提出建议。

一、背景介绍

随着社会的不断进步，手机行业得到了快速发展，人们的生活也逐渐被各种手机软件填满，手机游戏更是给人们的日常生活带来许多乐趣，人们利用手机游戏来打发空闲时间，增添了许多娱乐性和趣味性。与最初画面简陋、规则简单、玩法单一的手机游戏相比，如今的游戏已经经历了重大的变革，发展成为游戏性强、画面优美的优秀产品，可以和掌上游戏机媲美。随着互联网的快速发展，多人联网游戏出现，丰富了人们游戏的途径，比如近期的王者荣耀和吃鸡游戏，均受到众多用户的喜爱。

2017 年 8 月 25 日，全球领先的新经济行业数据挖掘和分析机构 iiMedia Research(艾媒咨询)权威发布《2017 上半年中国手机游戏市场研究报告》。报告中显示，2016 年中国手游用户规模达 5.23 亿人，市场规模 783.2 亿元，预计到 2017 年底，中国手游市场将突破千亿。



在众多的手机游戏中，哪些手机游戏深受大家的喜爱？事实上，手机游戏评分是多种因素综合作用的结果，到底有哪些因素在影响手机游戏评分呢？评分的差异又是如何产生的？本报告收集了 1140 个安卓手机游戏相关数据，对手机游戏评分的相关影响因素展开研究。

二、数据说明

本报告使用的是课程提供的安卓手机游戏相关数据，共 1140 条记录，除去评分值缺失的样本，共有 938 条记录。在数据预处理阶段，我们对评分变量进行分类，大于 7.4 为 1，其余为 0；最后更新时间以 2016 年 12 月 31 日为基准点，计算相距天数；将类别重新整合为三类：益智休闲、动作竞争和扮演模拟工具。益智休闲中包括棋牌游戏、音乐游戏、策略塔防和冒险解谜；动作竞争中包括射击游戏、格斗游戏、竞速游戏、体育运动、飞行游戏和动作竞争；扮演模拟工具包括角色扮演、模拟经营、养成游戏和游戏工具。此外，我们利用原始数据中开发商信息，加入开发数和是否含有“游戏”变量，分别记录该开发商共开发了几款游戏和名称中是否含有 Games；利用游戏名称信息，加入是否含有数据包变量，若游戏名称中含有“（含数据包）”，则取值为 1，否则为 0；利用语言信息，加入是否含有中文、是否含有英文、是否含有日文和是否含有韩文四个分类变量；利用资费信息，加入是否完全免费分类变量。

模型构建所用数据共包含 16 个变量，因变量为安卓手机游戏评分，其他变量为自变量。具体的变量说明如表 2-1 所示。

表 2-1: 数据变量说明表

变量类型		变量名	详细说明	取值范围	备注
因变量		评分	分类变量	0 或 1	
自变量	外部因素	评论数	单位：人次	[0, 96354]	
		喜欢数	单位：人次	[0, 13323]	
		热度	单位：°C	[28, 95]	
	内部因素	开发数	单位：个	[0, 38]	
		类别	定性变量：共 3 个水平	益智休闲、动作竞争和扮演模拟工具	
		相距天数	单位：天	[90, 2340]	基准日为 2016 年 12

					月 31 日
		游戏版本	定量变量	[0, 47]	
		支持系统	定量变量	[1. 5, 5]	
		是否包含“游戏”	分类变量	0 或 1	
		是否含有数据包	分类变量	0 或 1	
		是否中文	分类变量	0 或 1	
		是否英文	分类变量	0 或 1	
		是否日文	分类变量	0 或 1	
		是否韩文	分类变量	0 或 1	
		是否免费	分类变量	0 或 1	

三、描述分析

游戏评分一共 938 个值，跨度从 1 到 9.6，均值和中位数均为 6.6，其中愤怒的小鸟太空版无广告版和愤怒的小鸟太空版占据了前两把交椅。从图 1 中可以看到游戏评分的分布呈现出双峰的形态，第一个峰值大概在 5 左右，第二个峰值大概在 8 左右。

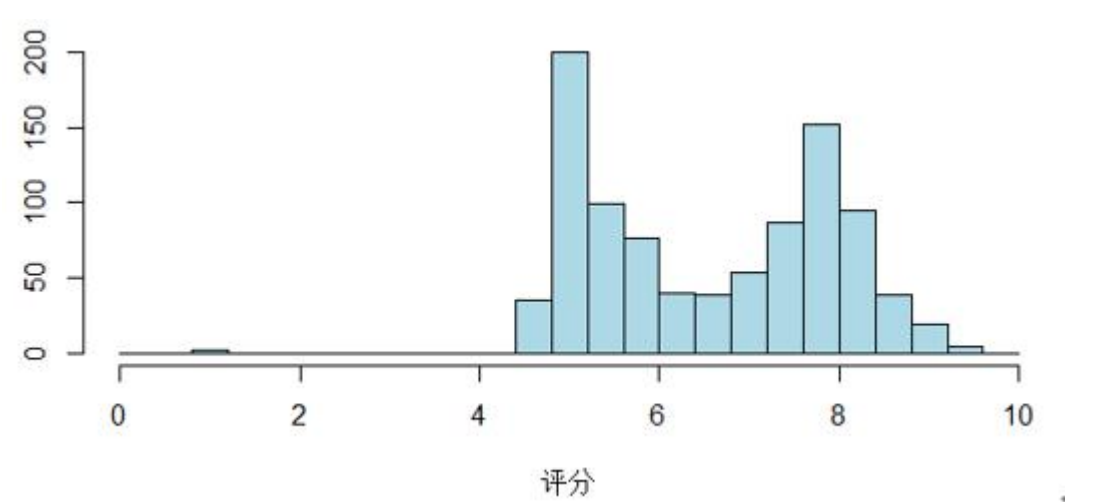


图 3.1 评分的直方图

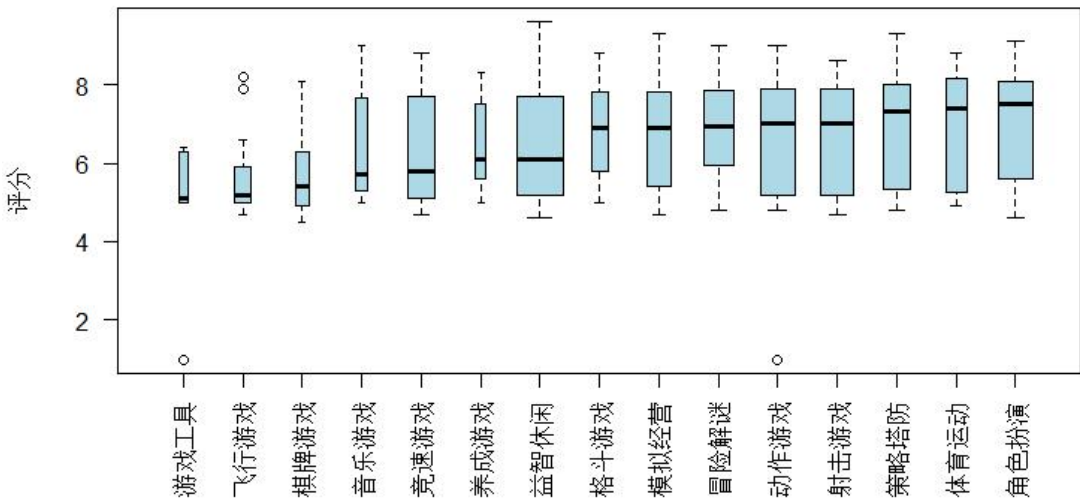


图 3.2 游戏类别与评分箱线图

从图 3.2 中可以看到对于不同类别的游戏，评分有着一定的差距。总体来说，游戏工具、飞行游戏和棋牌游戏的评分明显偏低，属于这些类别的游戏个数也比较少，而策略塔防、体育运动和角色扮演这些类别的游戏评分较高。

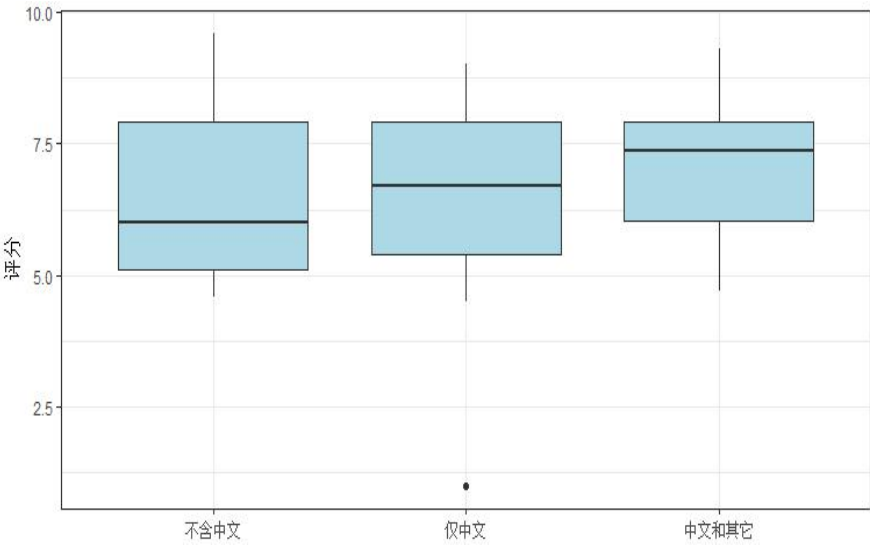


图 3.3 语言与评分箱线图

从图 3.3 中可以看到游戏的语言、不同对评分有者一定的影响，语言中包含有中文和其它语言的游戏评分总体上高于其它，仅含中文的游戏也比只有其它语言的游戏高。

从图 3.4 中可以看到，对于评分变量不同的样本，喜欢数和评论数的分布是有者明显差别的，评分变量为 0 的游戏喜欢数和评论数明显都集中在取值比较小的区域。当游戏的评论数小于 600 时，评分变量为 1 的概率非常小，但在该图中，喜欢数对 0, 1 评分变量影响并不明显。

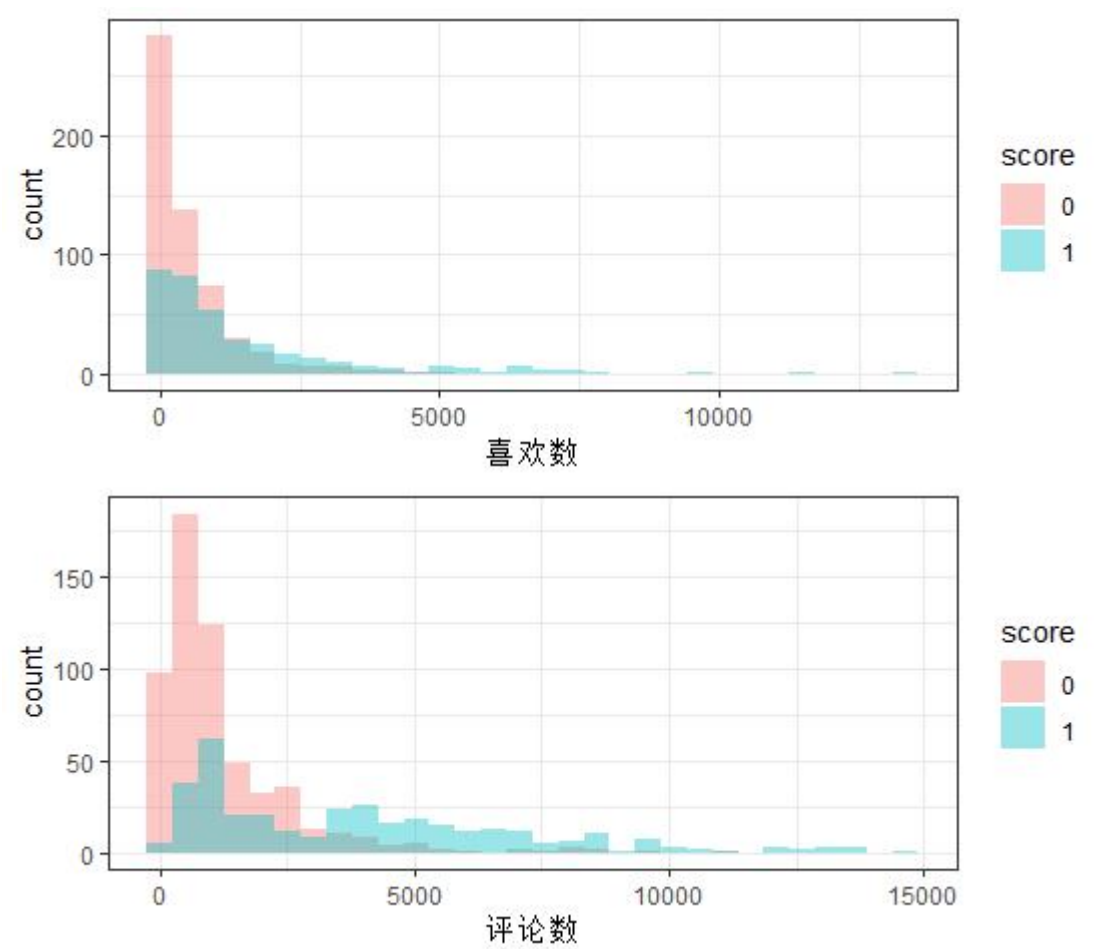


图 3.4 评论数和喜欢数在不同类别下的频次

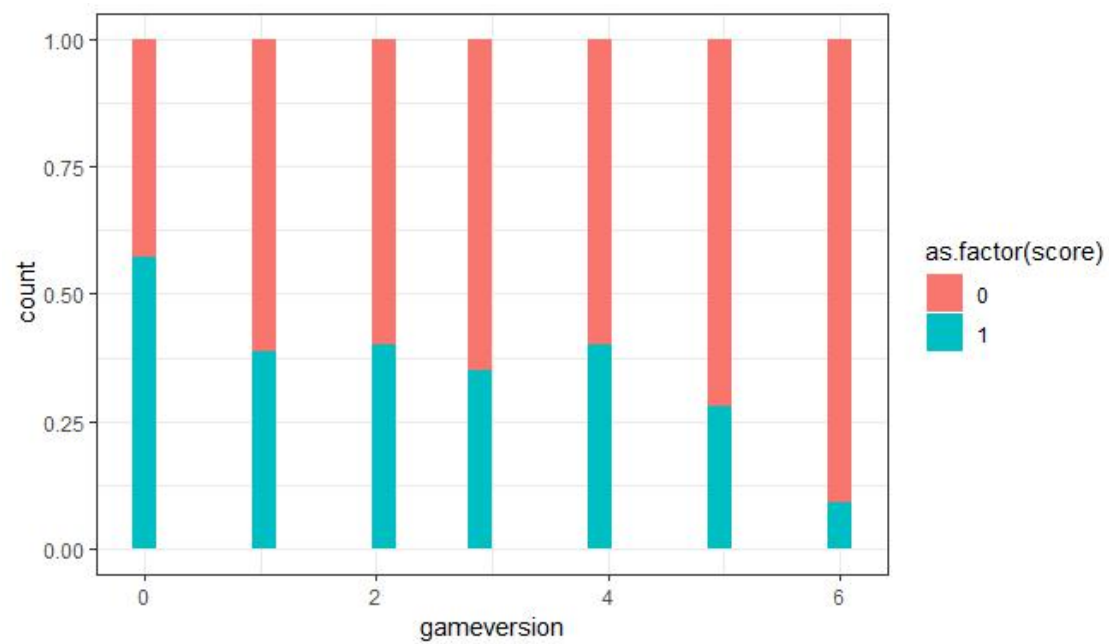


图 3.5 游戏版本在不同类别下的频次

从图 3.5 中可以看到，对于不同的游戏版本，评分变量取 0 和 1 的比例也有差异，游戏

版本为 5 和 6 的游戏评分取 1 的概率明显偏低。

四、模型建立

逻辑回归(LR)算法具有快捷、稳健、可解释性强的特点，是工业界最常用的模型之一。LR 模型对变量关系的线性限制，使得其难以达到最优。在建模的过程中，增加 L2 惩罚函数，减少过拟合。本文将 LR 模型作为基准，对数据预处理和模型表现作出快速评估。

(一) 多种分类算法综合比较分析

首先，我们将建立一个 logistic 回归模型来预测一款手机游戏的评分是否为 1。根据 15 个变量来判断手机游戏的评分高低。将手机游戏的历史数据作为逻辑回归的训练集，根据用户反馈信息和游戏特征属性建立一个分类模型，估计用户评分值。

为评估安卓手机游戏评分值，需要求解出 16 个参数 $\theta_0, \theta_1 \cdots \theta_{15}$ ，并根据所得参数值得到一个概率值。由概率值转化成 0,1 分类，这就是建立分类器的过程。在建立分类器之前，需设定阈值，根据阈值判断违约结果。若设定阈值为 0.5，游戏评分的概率值大于 0.5 被判定为高分游戏，反之，被判定为低分游戏。

在将清洗干净的数据代入分类器进行训练前，本文首先对数据进行正则化处理。采用 Standard Scaler 正则化去除手机游戏数据大量有偏分布和极端值，只保留数据的排序关系。正则化的过程是把数据集中的每个样本所有数值缩放到(-1,1)之间，将每个样本缩放到单位范数。

为减少过拟合，本文采用 5 折交叉验证法估计模型参数。先进行洗牌操作，将样本数据随机的切分成 5 份互不相交的大小相同的子集；每次取 1 份为测试集，其他 4 份为训练集。对此训练集进行训练，得到一个模型。依次取不同的 1 份为测试集，重复上述过程 5 次，可得 5 个模型。最终选出 5 次评测中平均测试误差最小的模型，将测试集数据代入该模型，得出预测值。

利用网格搜索法在训练集上进行超参数整定，调整或更改默认参数值以获得更好的模型。

准确率是预测正确的个数除以总样本个数，表明了预测值和真实值的差异，反映了分类器对整个样本的判定能力。在样本不均衡的时候，即使精度可以达到很高，但是模型没有用

处, 准确率具有欺骗性。因此, 本文又引入精确度和召回率来综合评估模型预测效果。精确度反映了被分类器判定的正例中真正的正例样本的比重。召回率反映了被正确判定的正例占总的正例的比重。

逻辑回归算法在不同阈值条件下的模型评估效果如图 4.1。

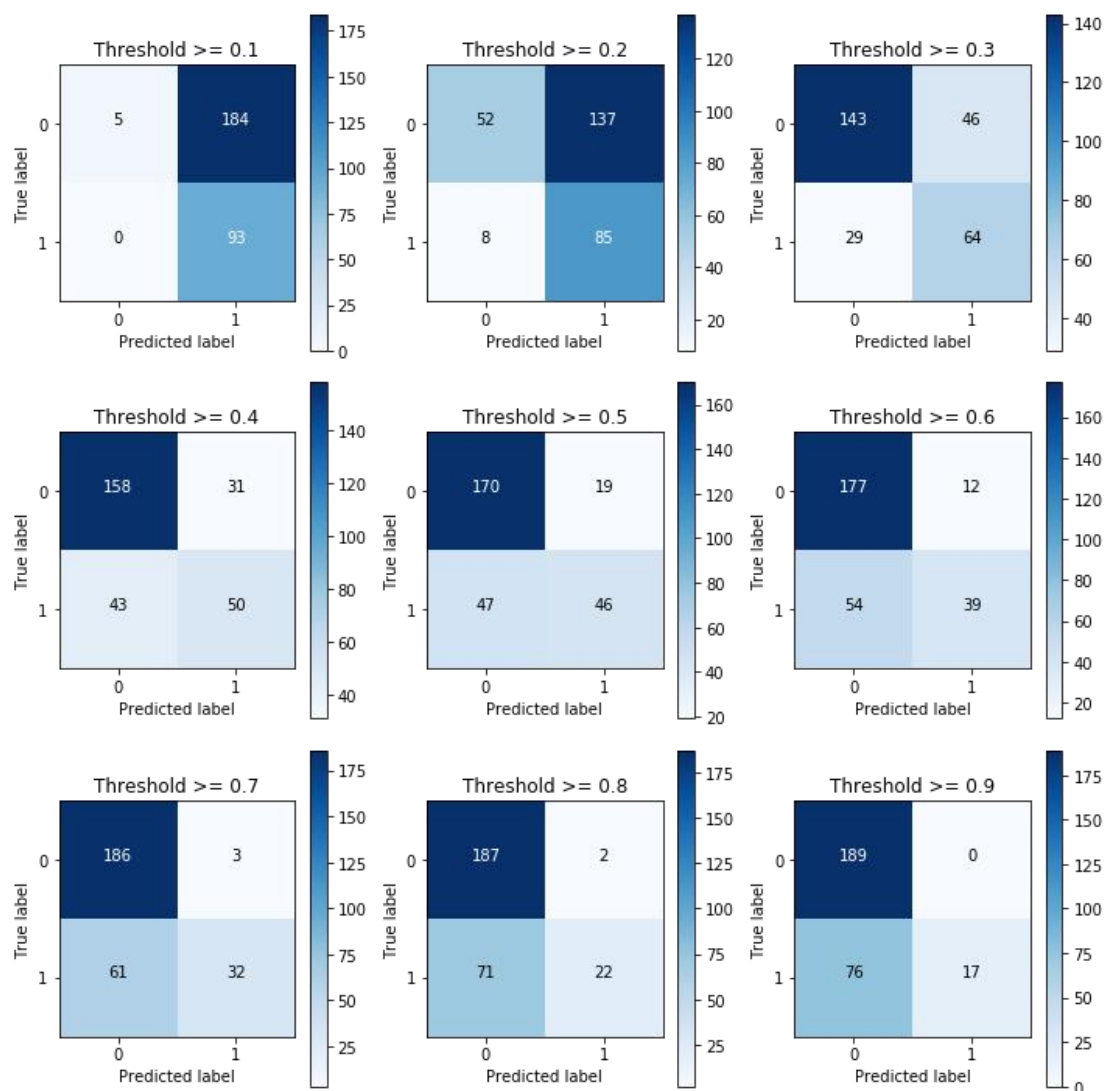


图 4.1 不同阈值下逻辑回归算法的分类效果

混淆矩阵中右上角是误杀值, 左上角和右下角为成功预测的样本个数。从图 4.1 可得, 逻辑回归阈值的设置对召回率有很大影响。阈值越大, 说明筛选越严格。设置阈值为 0.4 时, recall 值较高, 此时的精度也高。权衡召回率、精确度和准确度可以选取阈值为 0.4。

K-近邻(KNN)算法通过计算新数据与训练数据特征值之间的的距离, 选取 K 个距离最近的邻居进行分类, 新数据将被分配给其近邻的类。经过参数优化过程, 得出 K 为 5 时,

模型的预测误差最小，效果最好。

SVM 是一个强大的分类算法，模型的精准度与参数集有很大关系。当 C 很大时，虽然可以导致分错的点更少，但是过拟合情况会较严重。我们使用网格搜索法在训练集上寻找最好的参数，得出惩罚因子 C 为 0.5，核函数为线性核函数最佳。同理，我们得出决策树、朴素贝叶斯、线性判别分析和随机森林等分类算法的最优参数。

针对手机游戏数据，分别采用上述 7 种算法进行分类，给出不同模型下的结果对比，见表 4.1。

表 4.1 不同模型结果分析

分类器	准确率(%)	精确度(%)	召回率(%)
逻辑回归	80.26	72.31	62.67
支持向量机	82.83	85.71	56.00
决策树	73.39	60.00	52.00
KNN	81.55	79.63	57.33
朴素贝叶斯	79.40	75.47	53.33
线性判别分析	77.68	84.85	37.33
随机森林	80.69	74.14	57.33

由表 4.1 结果可得，支持向量机有最高的准确率和精确度，而逻辑回归有更高的召回率。权衡准确率、精确度和召回率三项评估指标，支持向量机的表现效果最好。

为了更直观展现不同算法 Recall 值的差别，图 4.2 给出混淆矩阵。

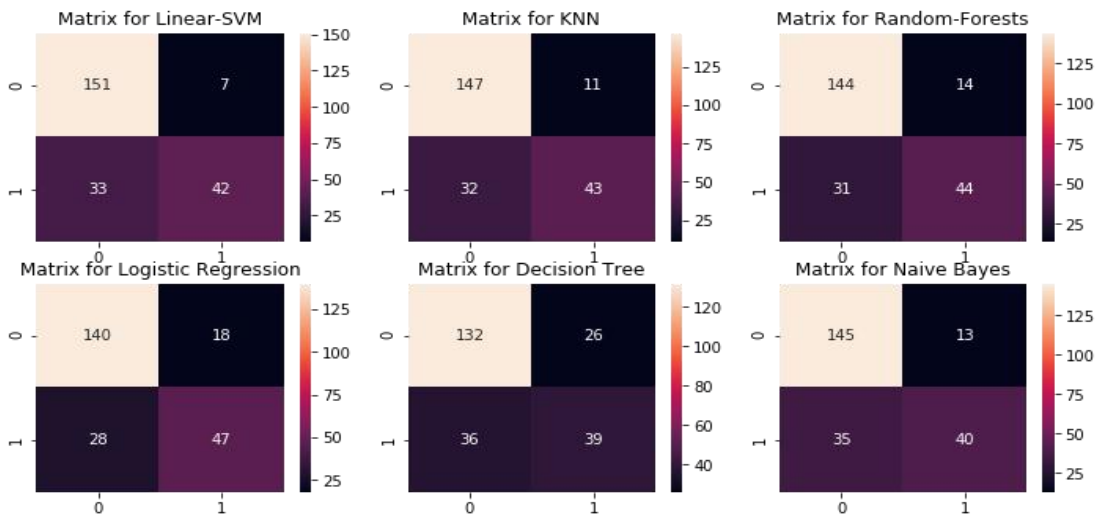


图 4.2 不同分类器结果比较

从图 4.2 可以看出逻辑回归的 Recall 值最大, L-SVM 其次。但 SVM 的精确度明显高于其他算法。最终, 我们选取 SVM 分类器作为最终模型。

(二) 集成算法提升模型精度

为了进一步提升模型的精度和性能, 本文选取投票分类器、Bagging 元估计器、Adaboost 和梯度树提升这四种集成算法训练分类模型。集成方法的目标是把多个使用给定学习算法构建的基估计器的预测结果结合起来, 从而获得比单个估计器更好的泛化能力(鲁棒性)。

本文中, 投票分类器结合了 KNN、随机森林、逻辑回归、决策树、朴素贝叶斯、支持向量机 6 个机器学习分类器, 利用平均预测概率(软投票)的方式来预测分类标签。 Bagging 元估计器中, 本文分别构造 KNN 和决策树估计器的 bagging 集成, 比较得 KNN 估计器的集成准确率更高. Adaboost 提升算法是一个逐步增强的弱模型, 弱学习或估计是一个决策树, 利用超参数整定得出最优参数, 学习率取 0.05, 树的个数为 200, 最终准确率为 84.16%。

将以上结果绘制为条形图, 如图 4.3。

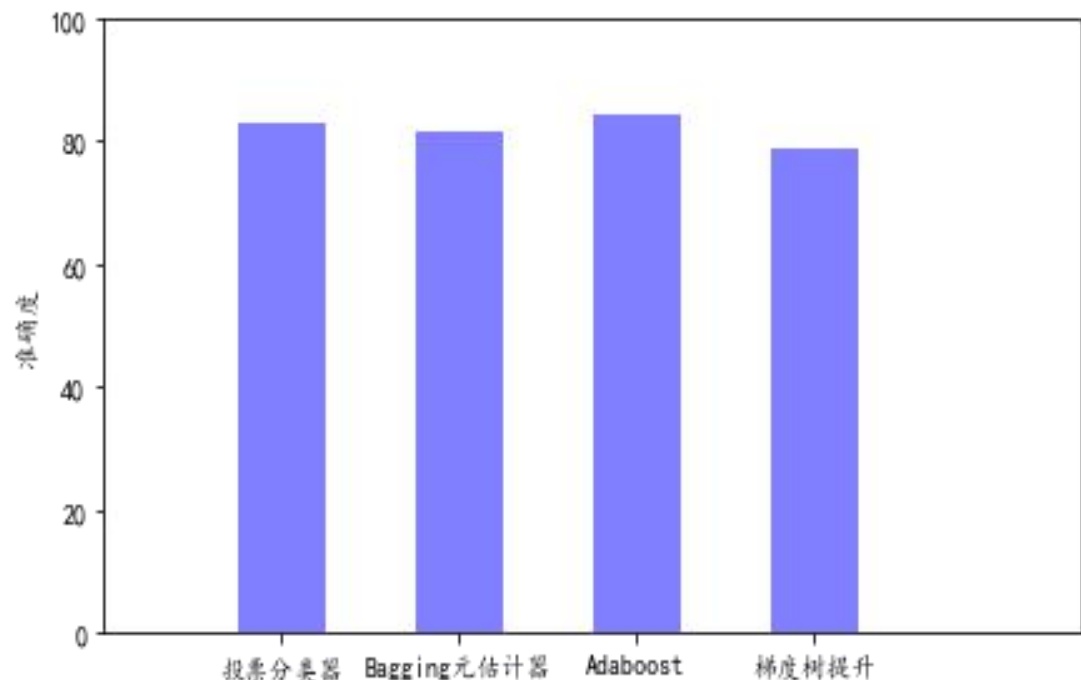


图 4.3 不同集成算法准确度比较

由图 4.3 可知, 四种集成算法的准确率均在 80%以上, AdaBoost 集成算法准确度最高,

投票分类器, Bagging 元估计和梯度树提升算法与 AdaBoos 算法的效果相差不大。综上所述, Adaboost 算法训练出最优分类器。

五 结论与建议

本案例对 2010 年 8 月 5 日-2016 年 10 月 2 日之间更新的部分安卓手机游戏数据进行统计分析, 以逻辑回归为基准模型, 采用 SVM 和决策树等 7 种分类器对手机游戏数据集进行分类。利用 Adaboost 等集成算法进一步得出准确率更高的模型, 最终建立 Adaboost 分类器。安卓手机应用商店可以根据已建立的模型, 对手机游戏评分进行分类。若预判手机游戏的评分较高, 可以在应用商店中引入此游戏, 反之, 则不引入此游戏。

利用随机逻辑回归进行特征筛选分析手机游戏评分影响因素, 得到如下结论:

(1) 影响手机游戏评分的主要因素有: (1) 外部因素: 评价数、喜欢数; (2) 内部因素: 游戏商开发游戏个数、游戏是否含数据包。

(2) 手机所属类别对评分值的影响有所区别。冒险解谜、策略塔防等益智休闲类手机游戏评分值较低, 动作、射击等动作竞争类手机游戏相比而言人气值较高。

手机游戏开发者可以根据以上影响因素, 完善所开发的游戏, 从而提高手机游戏的评分值。