

2018

美业自动化 策略更新报告

报告时间: 2018/6/27

不含生美/医美,金额大小,是否 LKA 商户的差异化策略

QINGQI SHI(石庆琦)

目录

I. 策略更新原因	2
II. 策略更新目标	2
III. 模型的初步筛选	2
IV. 模型的进一步筛选以及模型组合策略的选择	3
A. 评估不同的模型组合策略	3
B. 评估四个模型对 path_reject 的选择的有效性,以确定是否加入 posjxl20_m 和 cashl_m 用于 path_reject 的选择	6
C. 考虑对既没有 bt3_m 分数,又没有 cup_m 分数的申请单使用 posjxl20_m 模型	7
D. 根据 a),b),c)的分析,模型的进一步筛选以及模型组合策略的选择如下	8
V. 模型策略的参数(不同 path 的模型分阈值)选择	8
VI. 附录 1	10
VII. 附录 2	28
VIII. 附录 3	30

I. 策略更新原因

1. 由于聚焦建立3模型(PX5_UE_Model13/Model13PostScoreV2)需要用到芝麻分数据,无法继续使用,白户的自动化策略需要用其他模型
2. 美业项目申请量逐渐增加,信审压力增加,需要减少信审处理的申请单量
3. 美业自动化策略已较长时间未更新,可能需要使用近期数据重新评估

II. 策略更新目标

使用3到4个模型完成自动化策略,摆脱对芝麻分(马上芝麻分)的依赖,以调整自动化审批比例,同时保持通过率不降低,保持逾期指标基本不变差

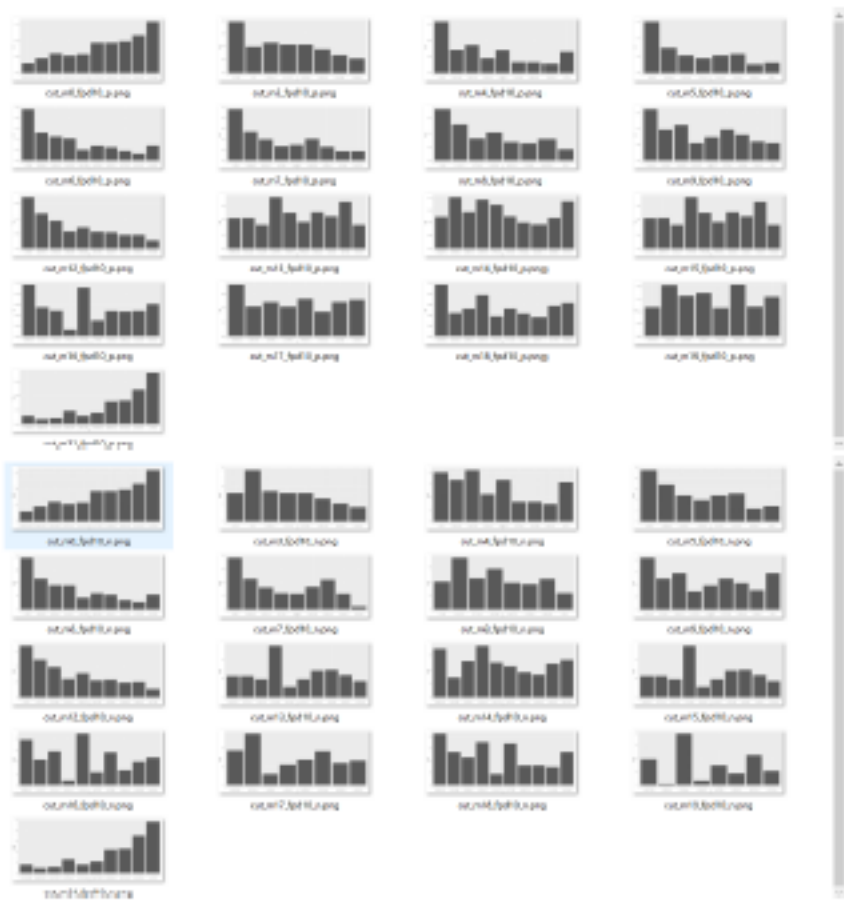
III. 模型的初步筛选

已接入的模型共有 27 个

将模型有效数量(具有有效模型分的申请单数量)低于 10000 的剔除,剩余 20 个模型,这 20 个模型在每天的覆盖情况如下(红色为未覆盖,蓝色为覆盖,横轴为 2017/4/16 到 2018/4/15)



剔除掉只有两个值的 3 个模型, 将剩余模型根据分数 10 等分, 画出每等分申请单中含有 FTD0 的申请单的比例或者数量, 17 个模型对 FTD0 的区分能力如下图所示



据此筛选出了对 FTD0 的区分能力较强的 8 个模型, 分别是

中文名	原文中的位置	英文命名(用于后续分析代码)
特征 X 0	"/Application/@tempVector0_r"	bt2_m
pboc 聚值立	"/Application/pboc_furini_model/@resModdb1"	phj1_m
现金贷模型	"/Application/DT_cashloan_model/@resModdb1"	caah1_m
商品贷聚值立 v20	"/Application/DT_M002L/@x1_pos_score_v20"	posjr120_m
商品贷聚值立 v21	"/Application/DT_M002L/@x1_pos_score_v21"	posjr121_m
分期模型	"/Application/Installment_Model/@resModdb1"	insg_m
聚值立	"/Application/POS_DT_Model3/@x1PosScoreV3"	js1_m
联合分	"/Application/G27/JointModelingScore/@data"	cup_m

对这 8 个模型进行进一步的分析, 如下图所示

B. 评估四个模型对 path_reject 的选择的有效性,以确定是否加入 posjxl20_m 和 cashl_m 用于 path_reject 的选择

图 1. 有征信客户 (b33 = 有值)

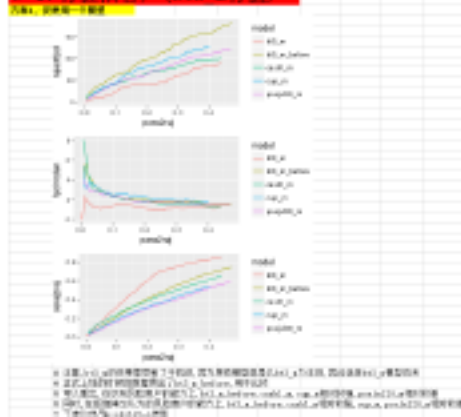
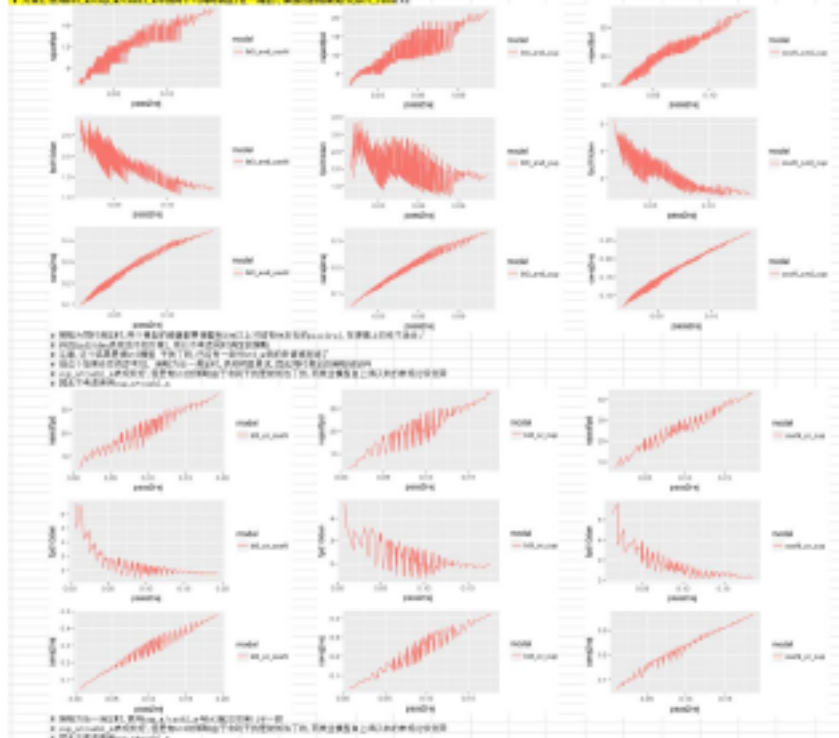
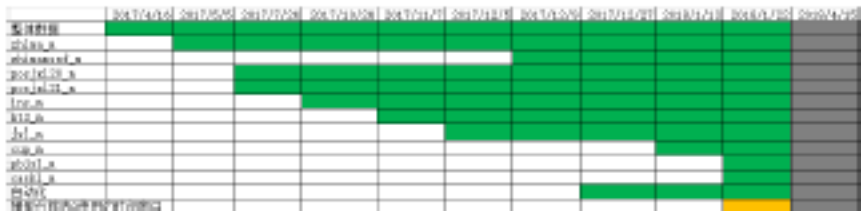


图 2. 有征信客户 (b33 = 有值)



IV. 模型的进一步筛选以及模型组合策略的选择

各个模型的接入时间图如下



因此我们用于策略选择的数据为, 2018/1/22 到 2018/4/15 这段时间的所有业务申请单

待选模型是否需要征信数据如下:

"bt3_m", "caxhi_m", "posjx120_m", "csp_m"

要征信 要征信 不要征信 不要征信

A. 评估不同的模型组合策略

我们有以下两个模型组合策略

组合策略 1: 有征信数据的选用一个模型, 无征信数据的选用一个模型, 根据模型分数高低划分为

path_gn, path_l, path_h 和 path_reject

组合策略 2: 有征信数据的选用两个模型, 共同决策, 无征信数据的选用一个模型, 根据模型分数高低划分为

path_gn, path_l, path_h 和 path_reject

通过编写一个模型函数, 并手动调整到大致效果不错的参数, 我们可以获取不同策略以及模型选择下的表现 (包括通过率, 人工比例和逾期表现等), 以此比较不同选择

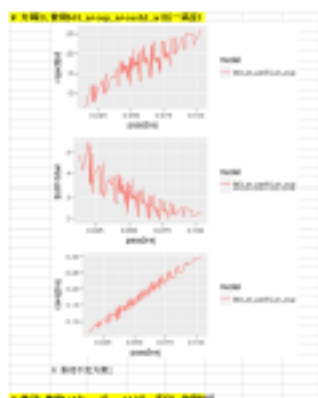
详见附录 1

结论:

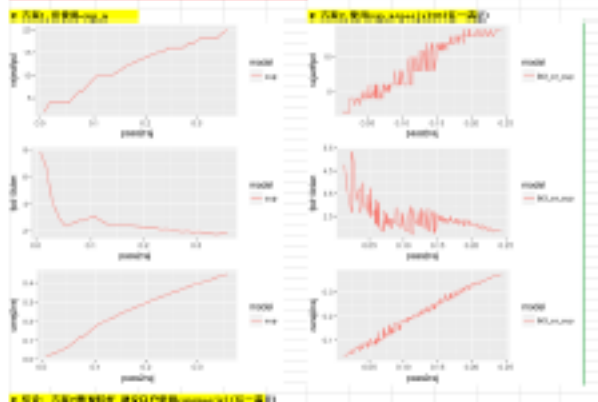
bt3_m 和 csp_m 整体表现较好

有征信数据的申请单, bt3_m 和 csp_m 叠加的表现较好

posjx120_m 和 caxhi_m 可以进一步提高



2. 无征信客户 (bt3_m 无值)



C. 考虑对既没有 bt3_m 分数,又没有 cup_m 分数的申请单使用 posjxl20_m 模型

仅有一单 fpxd10, 且处于 posjxl20 的高分, 因此 fpxd10 无法参考

有三单 maxdpd10, 分别处于 posjxl20 71%, 55%, 41% 分位, 可以参考

建议 57% 以上 continue

57% 以下 ce

10% 以下 reject

预计每天因此转 ce 的单数为每天 4.5 单

D. 根据 a),b),c)的分析, 模型的进一步筛选以及模型组合策略的选择如下

- i. 模型组合策略:
- # 策略:
- # 有征信用户 (model1~model2), 策略表如下
- | | | | | | | |
|---|----------|----------|----------|--------|--------|----------|
| # | # | | | | | |
| # | model1\2 | pass | continue | ce | reject | NA |
| # | # | | | | | |
| # | pass | pass | continue | ce | reject | pass |
| # | continue | continue | continue | ce | reject | continue |
| # | ce | ce | ce | ce | reject | ce |
| # | reject | reject | reject | reject | reject | reject |
| # | # | | | | | |
- # 无征信用户 (model3~model4), 策略表如下
- | | | | |
|---|----------|-------------|--------|
| # | # | | |
| # | model3\4 | notreject | reject |
| # | # | | |
| # | pass | pass | reject |
| # | continue | continue | reject |
| # | ce | ce | reject |
| # | reject | reject | reject |
| # | NA | ce/continue | reject |
| # | # | | |
- # 无征信用户且 model3 无值:ce
- ii. Model1~4 的选择
- Model1: ht3_m
- Model2: csp_m
- Model3: csp_m
- Model4: posx120_m

V. 模型策略的参数(不同 path 的模型分阈值)选择

根据以上模型组合策略, 编制了函数 get_new_table3 和 get_result3, 用于模型美业自动化策略划分 path_go, path1, path2 和 path_reject, 以及生成模型结果和各类指标, 函数介绍见附录 2

参数选择过程如下

1. 美业自动化策略更新的原因

1. 现有美业自动化策略使用了美业 3.0 模型(@tempVector6_r),聚信立 3.0 模型 (POS_JXL_Model3/@jxlPosScoreV3)和芝麻分 (extendSourceLists[@data_source_code="Zhima_score" and @param_name="score"]/@param_value)的数据,其中聚信立 3.0 模型也需要用到芝麻分(马上芝麻分)数据,为了摆脱对芝麻分的依赖,需要用其他模型替换聚信立 3.0 模型和芝麻分.
2. 美业项目的每日申请量逐渐增加,信审压力增加,需要减少信审处理的申请单量.近期信审处理的申请单占比约为 36%.
3. 美业自动化策略已较长时间未更新,客户资质分布可能发生了变化,需要使用近期数据重新评估.

2. 美业自动化策略更新的目标

计划使用 3 到 4 个模型完成新的自动化策略, 相比现有美业自动策略, 提高自动化审批比例,同时保持通过率不降低,保持逾期指标基本不变差.根据近期数据,美业项目自动化审批比例为 64%,通过率为 65%,fpd10 比例为 0.57%. 更新的目标为美业项目自动化审批比例提高到 72%(相应的, 信审处理的申请单占比降低到 28%),通过率不低于 65%,fpd10 比例不低于 0.60%.

3. 美业自动化策略更新展示

根据美业项目 2018/1/22 到 2018/4/15 的数据,利用 impala,RStudio server,excel 等分析工具,考虑不同模型的稳定性,有效性,经过模型的初步筛选,模型的进一步筛选以及模型组合策略的选择,模型策略的参数(不同 path 的模型分阈值)选择这三个步骤,更新后的的美业自动化策略展示如下:

我们将使用的模型包括美业 3.0 模型(@tempVector6_r),银联分模型 (CUPJointModelingScore/@data)和 R 聚信立模型(R_JXL_MODEL/@jxl_pos_score_v20).

对于有征信数据的用户,我们使用美业 3.0 和银联分模型共同决策,对于每个模型,我们根据模型分的好坏设置 3 个阈值,将其分为 pass,continue,ce,reject 四个分数段, 然后根据联合决策表进行路径决策.

美业 3.0 模型阈值如下,注意,美业 3.0 模型分数越低代表客户资质越好:

分数段	美业 3.0 模型分数	在样本美业 3.0 模型分中的分位数(越高资质越好)
低于此值 pass	0.019408524	46%
低于此值 continue	0.031456888	27%
低于此值 ce,高于此值 reject	0.09292867	4%

银联分模型阈值如下,注意,银联分模型分数越低代表客户资质越好:

分数段	银联分模型分数	在样本银联分模型分中的分位数(越高资质越好)
低于此值 pass	0.057681888	37%
低于此值 continue	0.116822641	11%
低于此值 ce,高于此值 reject	0.22404933	2%

联合决策表如下:

美业 3.0\银联分	pass	continue	ce	reject	NA
pass	path_go	path1	path2	path_reject	path_go
continue	path1	path1	path2	path_reject	path1
ce	path2	path2	path2	path_reject	path2
reject	path reject	path reject	path reject	path reject	path reject

旧策略的指标

[illegible]

圖書出版

small VOT*acc_change[1]:1,1[2]	
path	0.000
ca	80.100
cont_loss	80.220
mutual_info	80.220
total	80.220
small VOT*april:1,change[2]:1,1[2]	
path	0.000
ca	0.000
cont_loss	0.000
mutual_info	0.000
total	0.000

100

[illegible]

策略閾值篩選

模型	变量	系数	标准差	t 统计量	p 值
模型 1	控制变量	0.12	0.05	2.40	0.02
模型 2	控制变量 + 交互项	0.15	0.06	2.50	0.01
模型 3	控制变量 + 交互项 + 二次项	0.18	0.07	2.57	0.01

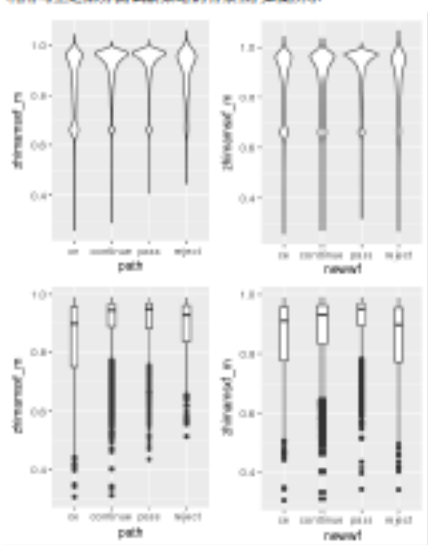
因此，`model` 和 `model_name` 的數量遠超過其他工業類別的數量，因此工業的 `model` 和 `model_name` 的標度

策略結果

	I	II	III
FAB	40	20	5
FAD	20	10	2
FAC	10	80	0
FAG	80	70	0

[illegible]

利用马上决策分割试新策略的有效性,如图所示



更详细的指标参见附录B。

对于无征信用户,我们使用银联分模型和 R 聚信立模型共同决策, 对于银联分模型, 我们根据模型分的好坏设置 3 个阈值,将其分为 pass,continue,ce,reject 四个分数段, 对于 R 聚信立模型, 我们根据模型分的好坏设置 3 个阈值,将其分为 continue, ce,reject1,reject2 四个分数段, 然后根据联合决策表进行路径决策。

银联分模型阈值如下,注意, 银联分模型分数越低代表客户资质越好:

分数段	银联分模型分数	在样本银联分模型分中的分位数(越高资质越好)
低于此值 pass	0.037972549	71%
低于此值 continue	0.051390819	45%
低于此值 ce,高于此值 reject	0.145530969	4%

R 聚信立模型阈值如下,注意, R 聚信立模型分数越高代表客户资质越好:

分数段	R 聚信立模型分数	在样本 R 聚信立模型分中的分位数(越高资质越好)
高于此值 continue	0.962759256	64%
高于此值 ce	0.907895029	14%
高于此值 reject1,低于此值 reject2	0.786736319	2%

联合决策表如下:

银联分\R 聚信立	continue	ce	reject1	reject2	NA
pass	path_go	path_go	path_go	path_reject	path_go
continue	path1	path1	path1	path_reject	path1
ce	path2	path2	path2	path_reject	path2
reject	reject	reject	reject	path_reject	reject
NA	path1	path2	path_reject	path_reject	path2

4. 美业自动化策略更新后的效果预计,数据为 2018/1/22 到 2018/4/15 的数据

主要指标变化:

指标	更新前	更新后
信审处理的申请单占比 (path2)	35.693%	28.269%
自动化审批比例 (除开 path2)	64.306%	71.730%
初审处理的申请单占比 (path1)	29.291%	22.577%
核心自动化审批比例 (除开 path1,path2)	35.015%	49.153%
通过率	65.308%	65.348%
fpd10 比例	0.5708%	0.5837%

更新前后决策路径的转移

旧路径\新路径	path_go	path1	path2-pass	path2-reject	path_reject
path_go	4071	1174	194	129	50
path1	4631	4149	1268	845	249
path2-pass	926	2181	4758	0	429
path2-reject	360	1062	0	2463	326
path_reject	0	0	0	0	563

有征信数据部分主要指标变化

指标	更新前	更新后
信审处理的申请单占比 (path2)	34.024%	23.586%
自动化审批比例 (除开 path2)	65.976%	76.414%
初审处理的申请单占比 (path1)	30.464%	25.196%
核心自动化审批比例 (除开 path1,path2)	35.512%	51.218%
通过率	75.861%	74.726%
fpd10 比例	0.5774%	0.5815%

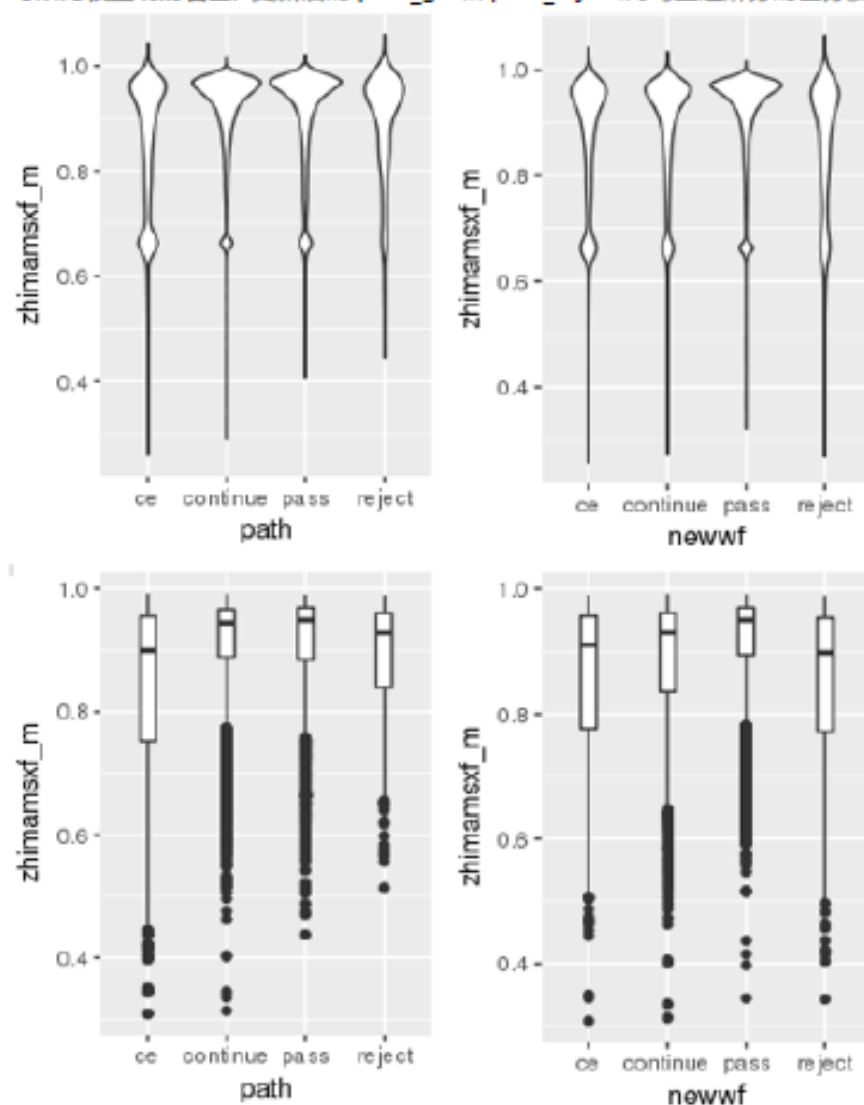
无征信数据部分主要指标变化

指标	更新前	更新后
信审处理的申请单占比 (path2)	38.942%	37.383%
自动化审批比例 (除开 path2)	60.244%	61.803%
初审处理的申请单占比 (path1)	27.008%	17.478%
核心自动化审批比例 (除开 path1,path2)	33.236%	44.325%
通过率	44.770%	47.098%
fpd10 比例	0.5489%	0.5806%

自动化策略路径决策对马上芝麻分的区分能力

纵轴为马上芝麻分,横轴为不同的决策路径,ce,continue,pass,reject 代表 path2,path1,path_go 和 path_reject,左侧为更新前,右侧为更新后.

可以比较直观的看出,更新后的 path_go 和 path_reject,对马上芝麻分的区分度更高了



5. 近期和未来美业自动化策略更新目标和计划

近期我们还将添加差异化的自动审批策略,对不同省份,不同商户类型(医美/生美),申请金额大小以及门店评级采用差异化的审批策略,以提高模型审批有效性.

未来我们将进一步减少 path_1 的数量以提高核心自动化审批比例.