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 | 5 |
| A. | 评估不同的模型组合策略 | 5 |
| | 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 | |
| CB | shl_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 | |
| VII. | 附录 2 | . 28 |
| VIIII | ut ⊕ a | 30 |

I. 策略更新原因

- 白干聚性立3模型(POS_IX_Model3/#jx1PosScoreV3)需要用到芝麻分数提。无故继续使用, 白产的自动化策略需要用 单纯维型
- 2. 美业项目申请量逐渐增加,信申压力增加,需要减少信申处理的申请单量
- 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)的按赖,以偶真自动化率挺比例。同时保持通过拿不降低。 保持逾期指标基本不变差

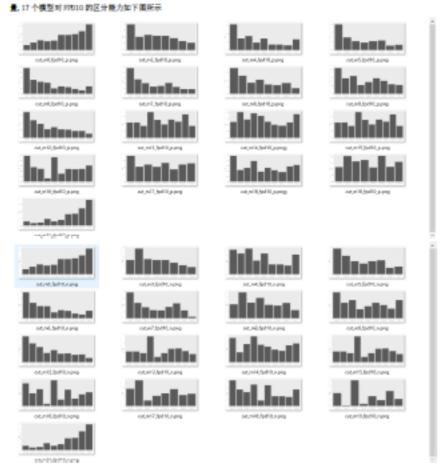
Ⅲ. 模型的初步筛选

已接入的模型共有37个

将模型有效数量(具有有效模型分的申请导数量)低于 10000 的制除, 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, 模粒为 2017/4/16 到 2018/4/15)



制除排尺有两个值的 3 个模型,将剩余模型模据分数 10 等分,面出每等分申请单中含有 27510 的申请单的比例或者数



据此筛选出了对 FF000 的区分能力较强的 8 个模型, 分别是

中文名。 报文中的位置。 英文命名(用于后楼分析代码)

美能工0, "/Application/WhempVectorをす", bt2.m

pboc 素値立、"/Application/pboc_juminli_model/@mes@oddbi",pbjml_m

現金貨模型 '/Application/DI_cashlosm_model/@resWoddbl', cashl_m

商品贷款债之 v20, '/Application/E_DL_MODL/@jrl_pom_score_v20', pomjrl20_m

商品贷款信文 v21, "/Application/生_DIL_MODIL/Birl_pos_score_v21", pos/st121_s

分解模型,"/Application/Installment_Wodel/像resWoddb1",ing.m

素体立、'/Application/POS_INL_Model3/@jxlPosScoreV3', jxl_m

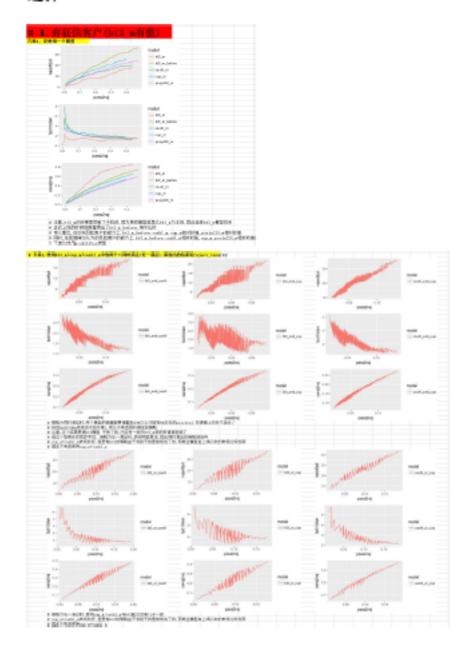
朝联行, '/Application/CDF Joint Modeling Score/ Modata', cup_ m

对这8个模型进行进一步的分析,如下国所示



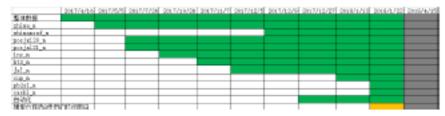
有如下结论

phjs1_s, posjs121_s, ins_s 这三个模型对 PRIO 的区分极力较强, 建议海汶. cashi_s 表现一般, bt1_s, posjs120_s, js1_s, cap_s 表现较好, 建议进一步得透 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



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费如下



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

待选模型技是否需要征信数据如下:

"bt3_m", 'cashl_m", 'posjx120_m", 'cap_m' 要任信 要任信 不要任信 不要任信

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

絕合策略 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, 根据模型分数真低划分为

path_go, path1, path2 #0 path_reject

組合策略 2: 有征性数据的选用两个个模型, 共同决策, 无征性数据的选用一个模型, 根据模型分数真低划分为

path_go, path1, path2 #0 path_reject

通过编写一个模拟函数。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,人工比例和逾期衰损等),以此比较不同选择

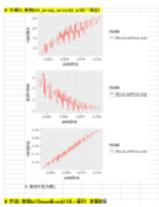
详见图录 1

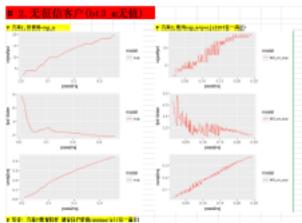
结论:

bt3_m 和 cup_m 整体衰疡较好

有征信数据的申请单。btil_m 和 cup_m叠如的表现较好

posjx120_m 和 cashi_m 可以返出情報





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 fpd10, 且处于 posjx120 的类分裂。因此 fpd10 无效参考 有三单 maxdpd10, 分别处于 posjx120 71%, 55%, 45%分裂。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

| 1. | 模型纸合策略: |
|-----|--|
| | ■ 景略: |
| | ■ 有征使用户(model1)model2)。景略表如下 |
| | # |
| | # model1\2 page continue ce reject NA |
| | s |
| | # pass pass continue ce reject pass |
| | # continue continue ce reject continue |
| | # ce ce ce reject ce |
| | # reject reject reject reject reject |
| | • |
| | ■ 无征信用户(model3=model4),景略表如下 |
| | • |
| | # model3/4 notreject reject |
| | • |
| | # pass pass reject |
| | # continue continue reject |
| | # ce ce reject |
| | # reject reject |
| | # NA ce/continue reject |
| | |
| | ■ 无征慎用产且 model3 无值:ce |
| 11. | Nodel1 ⁻⁴ 的选择 |
| | Nodel1: bt3_m |
| | Model2: cup_m |
| | Model3: cup_m |
| | Model4: posjx120_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,用于模型类型自动化策略划分 path_go, path1, path 2 和 path_reject,以及生成模型结果和各类指标,函数介绍见图录 2 参数选择过程如下

1. 吴业目动化策略更新的原因

- 現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
 (extendSoureceLists[@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.
- 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:

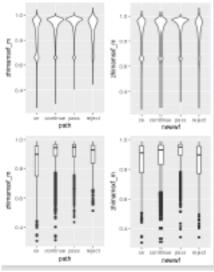
| 10 to 00 to 00 to 10 to 11 to 10 to | and the same of th | | | | | | | |
|---|--|------------------------|--|--|--|--|--|--|
| 分数段 | 银联分模型分数 | 在样本银联分模型分中的分位数(越高资质越好) | | | | | | |
| 低于此值 pass | 0.057681888 | 37% | | | | | | |
| 低于此值 continue | 0.116822641 | 11% | | | | | | |
| 低于此值 ce,高于此值 reject | 0.22404933 | 2% | | | | | | |

联合决策表如下:

| 4000000 | PH 1 - | | | | |
|-----------|-------------|-------------|-------------|-------------|-------------|
| 美业 3.0\银联 | pass | continue | ce | reject | NA |
| 分 | | | | | |
| 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 |

| | 的指标 | | | | | | | | | | | |
|--|--|--|-------------------------------|----------------------------------|-----------------------------|--|--|---|----------------------------|------------------|------------------------|--------|
| | teoer/TII.ttpl | | peng MCC ecube | | | peyd1E1cude | | | | | | |
| wt. | 1100 | | 14.77 | -1864 | | 0477 | -084 | | | | | |
| M | 70.008 | | 08 | 80,148 | | 08 | 8,37 | | | | | |
| reded. | 90.38.6 | | vent lease | 10.29 | | model. | N. 318 | | | | | |
| continue | 96.128 | | and conden | 16,236 | | work loose | 94, 136 | | | | | |
| ni centro | 50, 246 70, 646 | | ented ented still field by | 66,779 | | and a made | 8,18 | | | | | |
| total | Johnson TH. 147 | | march | State Chiefe | | record t E " todi it. | | | | | | |
| and a | dealer of a | | 19.1991 | 0.08 | | puth | dealer of the | | | | | |
| 1,000 | 0.78 | | confirmation | 0.6% | | 10, 2811 | 5.796 | | | | | |
| orien per | 0.408 | | model poor | 0.09 | | continue pare | 1, 194 | | | | | |
| old per | 0.08 | | 247 | 0.559 | | polel part | 5,275 | | | | | |
| 47 | 0.59 | | | 0.55 | | Total Control | 2.55 | | | | | |
| | 0.00 | | | | | | - | | | | | |
| No who date | 值筛选 | | | | | | | | | | | |
| 非(阿)阿 | 301 Mp 1/2 | | | | | | | | | | | |
| entrale la fa | | | | | | | | | | | | |
| Street, J.C. | \$ 20 M. Con. Co. | 1800年後年7日7 | 3,8 | | | | | | | | | |
| | 1 | P | - | | | | | | | | | |
| <u> </u> | | ACH | (0) | | | | | | | | | |
| | 10'M | ALC: U | 1CA | | | | | | | | | |
| Microsoph (C | CANAL COLUMN | CONSTRUCTION IN | 100 | | | | | | | | | |
| ica. | 12.70 | Self-er | 10 | | | | | | | | | |
| 14 | 757 | 2.5 | | | | | | | | | | |
| THE RESERVE | Laure Month on | PERMIT | TRIMERICAL | DECEMBER OF | of and the RM | 10 | | | | | | |
| | | | | | | | | | | | | |
| Morale Alb | JII. | | | | | | | | | | | |
| | 280 | | | | | | | | | | | |
| 東町箱 | | ORDIC STORAGE | | | | | | | | | | |
| 1000上年3046 | CAT IN THE REST | | | | | | | | | | | |
| 10年0.上年3544 | OR ALL PROPERTY | | | | | | | | | | | |
| iiiiの上東東iski mastalieris ii | I I | - 1 | | | | | | | | | | |
| 職の上等更終 soutabely fi eM | - | - | | | | | | | | | | |
| HEAD TO ME STAND CONTRACTOR OF THE STAND CARD | 1 | 1 | | | | | | | | | | |
| HIRO LEN EURO March Clerks III HIR HIR HIRO | - | 1 | | | | | | | | | | |
| HEAD TO ME STAND CONTRACTOR OF THE STAND CARD | | | | | | | | | | | | |
| IND. EN BOAR CONTROL OF TO B CASE CASE CASE | | 1 | Sarrie A, St | | | | | | | | | |
| Harris De Brands (1994) Salah Brands (1994) Salah Brands (1994) Salah Brands (1994) Salah Brands (1994) | | 1 1 2 2 3 4 4 5 4 5 6 7 6 7 7 8 7 8 7 8 7 8 7 8 7 8 7 8 7 8 | Canada A, Etc | | | | | | | | | |
| MARCHANIA MARCHA | | 3 3 8 8 8 8 8 8 8 | - Land 4, 51 | 10.700 | mar. è | tool set of | sold or | annie v | total e | as tur- | on tob | met. * |
| SECTION OF THE SECTIO | | 1 3 4 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 | | H-28 T-49 | 040.A | 100f_2CL/3: | 10.00 M | AND DESCRIPTION OF THE PERSON | | | | |
| SECULAR SOLITORS | | 50 51 51 51 51 51 51 51 51 51 51 51 51 51 | 011.108 17.108 | 9.36 7.10 | 01.07R -1.49 | Mod_ScL/h | 10/KL Nr 15.1% | 7.10 | 03.105 | 0.58 | 5.7% | 1.7 |
| GROUP A STANDARD OF THE STANDA | 1 2 3 3 5 6 6 6 7 8 7 8 7 8 7 8 7 8 8 7 8 8 7 8 8 1 8 1 | 50, 176 | 73,200 | 72-101 | -1,45 | 36.18 | 15.7% | 7.10 | 03.105 | 0.58 | 5.7% | 1.7 |
| HEAD OF STATE OF STAT | 1 2 3 3 5 6 6 6 7 8 7 8 7 8 7 8 7 8 8 7 8 8 7 8 8 1 8 1 | 50, 176 | 73,200 | 72-101 | -1,45 | 36.18 | 15.7% | 7.10 | 03.105 | 0.58 | 5.7% | 1.7 |
| SOLD SERVICES | | 2004 2004 2004 2004 2004 2004 | | 10.28 TL-10 10.28 TL-10 | 0.87.8 -0.87.8 -11.48 | Mod_SCL.05. 36.18 Mod_SCL.05. 36.79 | 10 (K), or 15, 178 16 (K), or 16, 178 | | 03.105 | 0.58 | 5.7% | 1.7 |
| this this this this this this this this | 1 2 3 3 5 6 6 6 7 8 7 8 7 8 7 8 7 8 8 7 8 8 7 8 8 1 8 1 | 50, 176 | 73,200 | 72-101 | -1,45 | 36.18 | 15.7% | 7.10 | 11.100 14.100 14.100 | 60,5000 0.338 | 0.7% 06,300 1.88 | 101.3 |

利用马上芝麻分割试新策略的有效性。如圆所示



更详细的指标表见图录 3

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 | 0.786736319 | 2% | | | | |
| reject2 | | | | | | |

联合决策表如下:

| 银联分\R 聚信 | continue | ce | reject1 | reject2 | NA |
|----------|----------|---------|-------------|-------------|---------|
| 立 | | | | | |
| 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 | 35.693% | 28.269% |
| (path2) | | |
| 自动化审批比例 | 64.306% | 71.730% |
| (除开 path2) | | |
| 初审处理的申请单占比 | 29.291% | 22.577% |
| (path1) | | |
| 核心自动化审批比例 | 35.015% | 49.153% |
| (除开 path1,path2) | | |
| 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 | 34.024% | 23.586% |
| (path2) | | |
| 自动化审批比例 | 65.976% | 76.414% |
| (除开 path2) | | |
| 初审处理的申请单占比 | 30.464% | 25.196% |
| (path1) | | |
| 核心自动化审批比例 | 35.512% | 51.218% |
| (除开 path1,path2) | | |
| 通过率 | 75.861% | 74.726% |
| fpd10 比例 | 0.5774% | 0.5815% |
| | | |

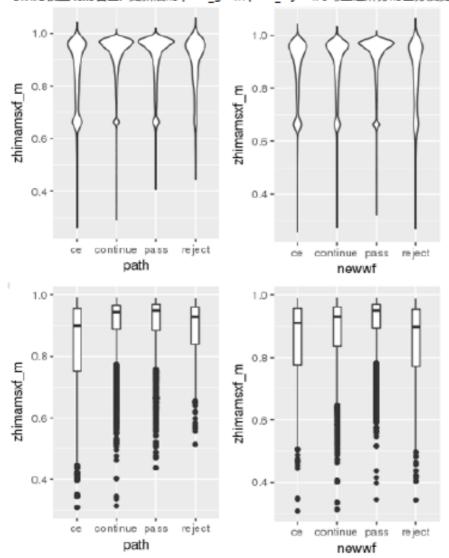
无征信数据部分主要指标变化

| 指标 | 更新前 | 更新后 |
|------------------|---------|---------|
| 信审处理的申请单占比 | 38.942% | 37.383% |
| (path2) | | |
| 自动化审批比例 | 60.244% | 61.803% |
| (除开 path2) | | |
| 初审处理的申请单占比 | 27.008% | 17.478% |
| (path1) | | |
| 核心自动化审批比例 | 33.236% | 44.325% |
| (除开 path1,path2) | | |
| 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.