



上海财经大学
Shanghai University of Finance and Economics



网约车国际化乘客应答取消问题实证分析

开题答辩

指导老师：王绍立 答辩人：何裕航



目录

CONTENTS



上海财经大学
Shanghai University of Finance and Economics



1

选题背景

2

国内外研究现状

3

研究的基本思路和方法

4

论文的创新及预期

5

论文安排及目录

PART ONE

01

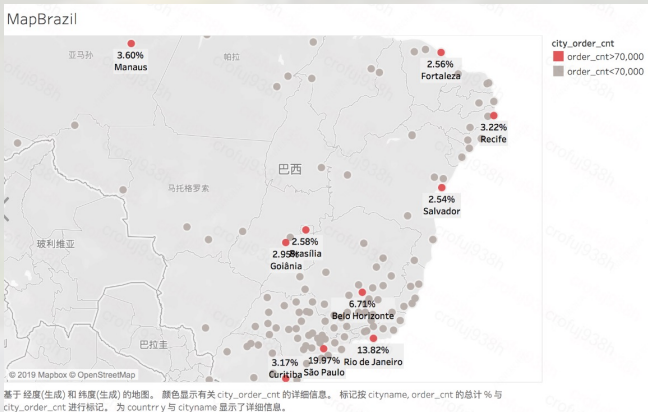
课题背景



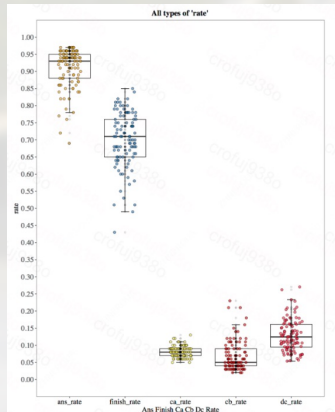
上海财经大学
Shanghai University of Finance and Economics

选题背景:

- 近年国内网约车行业经历了洗牌、整合和淘汰战，逐渐形成了相对稳定的格局。
- 目前在国内网约车渗透率极高的市场形势下，出于对市场增量的探索，迫使头部网约车平台的竞争向海外市场转移。其中巴西作为一个主要的市场探索方向。
- 在实际制定国际化市场运营方式时，在策略落地、客户端打签及精准推送、功能展示方式等环节，出现了运营方式战略不笃定、各个国家之间经验策略无法进行参照沉淀、各国政策不同的情况。
- 从分析实际乘客行为数据来看，巴西网约车市场虽然发展速度较快，但是仍然存在许多问题取消率较高，乘车安全风险问题等等需要研究。



通过分析巴西市场订单的地理分布，我们可以看出，有 61.12% 的订单集中在前 10 个城市中，其中圣保罗占到全国订单的 19.97%，里约热内卢占到全国订单的 13.82%，是非常依赖单体城市发展的市场。这其中当然也包含巴西许多城市的自身原因，例如争议地区或者危险地区等等，使得城市呈现头部集中度高的特征。



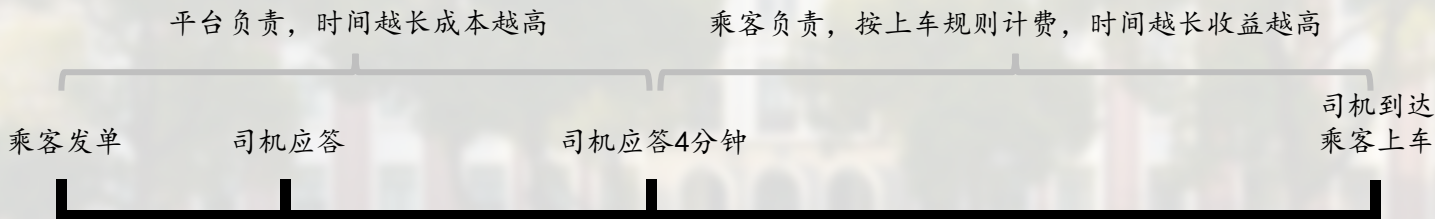
从订单的分城市核心指标分布中可以看到，除了各城市的完单率（finish_rate）之外，其他指标分布相对集中。对于巴西大部分城市，应答率（ans_rate）基本稳定于 0.9 以上。其中应答后乘客取消率（ca_rate）也很集中，基本上全部在 [0.05, 0.11] 的区间内。而完单率横跨范围最大，从 0.5 到 0.85 的城市均有。完单率的影响主要受取消率影响，明显看出有少量城市应答前取消率（cb_rate）过高（>0.15），需要进一步研究存在哪些問題。从各城市的乘客应答前、应答后取消率可以粗略看出，二者似乎存在一种动态平衡关系：乘客应答前取消率与乘客应答后取消率之和在 0.13 附近波动。

问题描述：

乘客应答取消问题并不适合直接划分为传统机器学习中的分类问题进行建模。理论原因如下：

- 1) 乘客应答取消行为与等待应答时间有密切联系，简单分类问题无法将概率与事件进行联系；
- 2) 该问题更关注条件概率，即当前乘客在已经等待 x 秒之后，取消订单的条件概率；
- 3) 数据中存在大量删失数据，即在乘客选择取消订单之前，平台就已经为其派单，或由金刚位转换到其他业务线应答。在分类问题下表现为，同样的特征下不同等待时间对应的标签不同。

巴西业务规则：



研究目的：以利益最大化为导向，

- 针对4分钟时间点采用的不同期望，采用不同的补贴经营策略应对：
- 4分钟前，识别高风险取消乘客特征，对其中高取消概率的用户，尽可能快速让乘客取消或确保等待至4分钟。
- 4分钟后，尽量不让乘客取消，持续贡献收益。关注乘客在应答前取消的具体时间，得到乘客的应答前取消概率。

选题意义：研究成果有相当强的现实意义，可以直接服务收益最大化。拓展到如差异化补贴、车型推荐、优化用户画像等业务场景，如：

1) 通过计算乘客在呼叫后发生应答取消的风险函数，可以帮助运营制定差异化补贴策略，节省补贴成本，提高收益。

现在网约车平台的补贴策略师根据乘客的活跃程度不同，发放不同面额的优惠券，使用短信或APP触达。因为没有考虑到乘客的出行需求、乘客画像等。这种简单粗暴的补贴导致券包核销率偏低，没有给平台带来更多订单量的同时还浪费用户触达的预算成本。本文通过计算得到乘客应答取消的风险函数和取消时长之后，运营可以在乘客发出需求之后，发送券包等激活乘客，促成完单率提升。

2) 有助于优化车型算法。当乘客发出出行需求以后，平台会展示车型选择界面，目前平台一般会根据价格高低将较低的两种车型默认选上，这样的做法没有考虑到乘客的车型偏好、历史叫车的行为数据，往往需要乘客结合自身出行需求进行修改后才最终呼出订单，影响乘客的出行体验。如果把风险函数作为乘客特征之一，对耐心度较高的乘客，可以匹配价格较低的但等待时间较长的车型，如此可以提高乘客的出行性价比体验。

3) 优化用户画像，给产品和运营方制定定制化出行方案提供更多可能。目前预测采用的方法是二分类机器学习法，即设定多个时间点，如呼叫10分钟后，通过二分类的机器学习方法进行取消概率预测，或者将经过集成方法进行特征重要性筛选后的变量输入至比例风险模型。二分类的做法对从业经验具有依赖性且没有考虑删失情况。本文基于真实数据集，摆脱Cox比例风险模型的假设限制，将基于深度学习的生存分析模型应用之场景，能够提高准确率，运营方可以基于此进一步定制个性化的出行方案。

PART TWO

02

国内外研究现状



上海财经大学
Shanghai University of Finance and Economics

在阅读相关文献时，主要从以下三个方面入手：

网约车相关

近年来，关于网约车平台相关的文献研究增多，但较多为基于定价策略研究，多为对实际将时间作为特征优化的研究较少。

生存分析、Logistic 方法论相关

生存分析的目的是估计直到感兴趣的特定事件发生的时间。本次准备使用的生存分析方法，文献方向多运用在较长的时间周期内，如医学观察，客户流失等，在网约车乘客行为方向应用文献较少。

Logistic方法在该领域的运用广泛。

联合建模相关

多为实证分析类论文，帮助就具体问题采用不同建模方法时使用。

近年来，关于网约车平台相关的文献研究增多，但较多为基于定价策略研究，多为对实际将时间作为特征优化的研究较少。主要阅读文献列表如下：

针对网约车乘客行为问题研究列举：

- [1] 郭敏，考虑乘客取消订单的网约车平台定价策略[J]。——以平台利润最大化为目标，建立乘客选择模型计算乘客取消订单概率，但仍使用二分类方法处理时间问题。
- [2] 王丹，考虑乘客应答单取消行为的网约车平台派单策略研究[D]。——考虑了出租车影响的应答单乘客取消模型，设计了司乘匹配动态仿真系统，但并未考虑来自乘客本身的影响。

针对其他双边平台领域用户行为问题研究列举：

- [3] 徐峰. 基于深度生存分析的电信客户流失研究[J], 经济与管理科学企业经济, 2022.
- [4] 孙凯. 基于组合预测算法的电音客户流失预警模型研究[D]. 云南财经大学, 2020.
- [5] 孙丽. 基于生存分析的A公司宽带客户流失问题研究[D]. 燕山大学, 2019.
- [6] 王钊, 蒋翠清, 丁勇. 基于混合生存分析的动态信信用评分方法[J]. 系统工程理论与实践. 2019

生存分析的目的是估计直到感兴趣的特定事件发生的时间。本次准备使用的生存分析方法，文献方向多运用在较长的时间周期内，如医学观察，客户流失等，在网约车乘客订单行为方向应用文献较少。Logistic方法应用文献较多。

针对生存分析的理论研究：

- [7] Cox, D. R. 1992. Regression models and life-tables. In Breakthroughs in Statistics. Springer. ——Cox DR在1972年第一次提出一种生存分析回归模型，被称为Cox回归模型。
- [8] Lee, E. T., and Wang, J. 2003. Statistical methods for survival data analysis, volume 476. John Wiley & Sons. ——Lee等人开始对感兴趣事件发生的时间信息进行观察。但是由于各种不同的原因，很多事件在观测期间内会失去跟踪，这就会使得数据发生删失。
- [9] Wang, Y.; Ren, K.; Zhang, W.; Wang, J.; and Yu, Y. 2016. Functional bid landscape forecasting for display advertising in ECML-PKDD.
- [10] 钱俊.生存分析中删失数据比例对Cox回归模型影响的研究[D]——针对删失数据比例对精确性的研究，在本次试验数据量下可认为此比例的生存分析结果依旧具有高可信性。
- [11] Jared Katzman, 2016, DeepSurv: Personalized Treatment Recommender System Using A Cox Proportional Hazards Deep Neural Network[J], Cornell University.
- [12] Hateftabar Fahimeh. The impact of psychological distance on tourists' length of stay: Survival analysis[J]. Journal of Hospitality and Tourism Management,2021,46.
- [13] Kaplan, E. L., and Meier, P. 1958. Nonparametric estimation from incomplete observations. Journal of the American statistical association.
- [14] Tibshirani, R. 1997. The lasso method for variable selection in the cox model. Statistics in medicine 16(4):385-395.
- [15] Faraggi, D. , & Simon, R. .(1995). A neural network model for survival data. Statistics in Medicine, 14(1),73-82.

针对生存分析的实证研究：

[16] Luck, M.; Sylvain, T.; Cardinal, H.; Lodi, A.; and Bengio, Y. 2017. Deep learning for patient-specific kidney graft survival analysis. arXiv preprint arXiv:1705.10245.——Luck等人把生存分析运用到医学研究中的临床分析，用疾病的发生作为事件并预测患者的生存时间。

[17] Jing. H., and Smola, A. J. 2017. Neural survival recommender. In WSDM, 515-524. ACM.——Jing等人提出了信息系统中的客户寿命估计，用来估计客户下次访问的时间。

联合建模相关论文，多为实证分析类论文，帮助就具体问题采用不同建模方法时使用。

针对实证分析联合建模的学术研究：

[18] 于铃玉. 基于反应和反应时间数据的联合模型[D]，经济与管理科学企业经济, 2022.

[19] 胡亚南. 基于纵向数据与生存时间数据联合建模的变量选择[J]，数理统计与管理, 2018.

[20] 带潜变量的区间删失数据的两种联合建模方法研究[D]，长春工业大学, 2018.

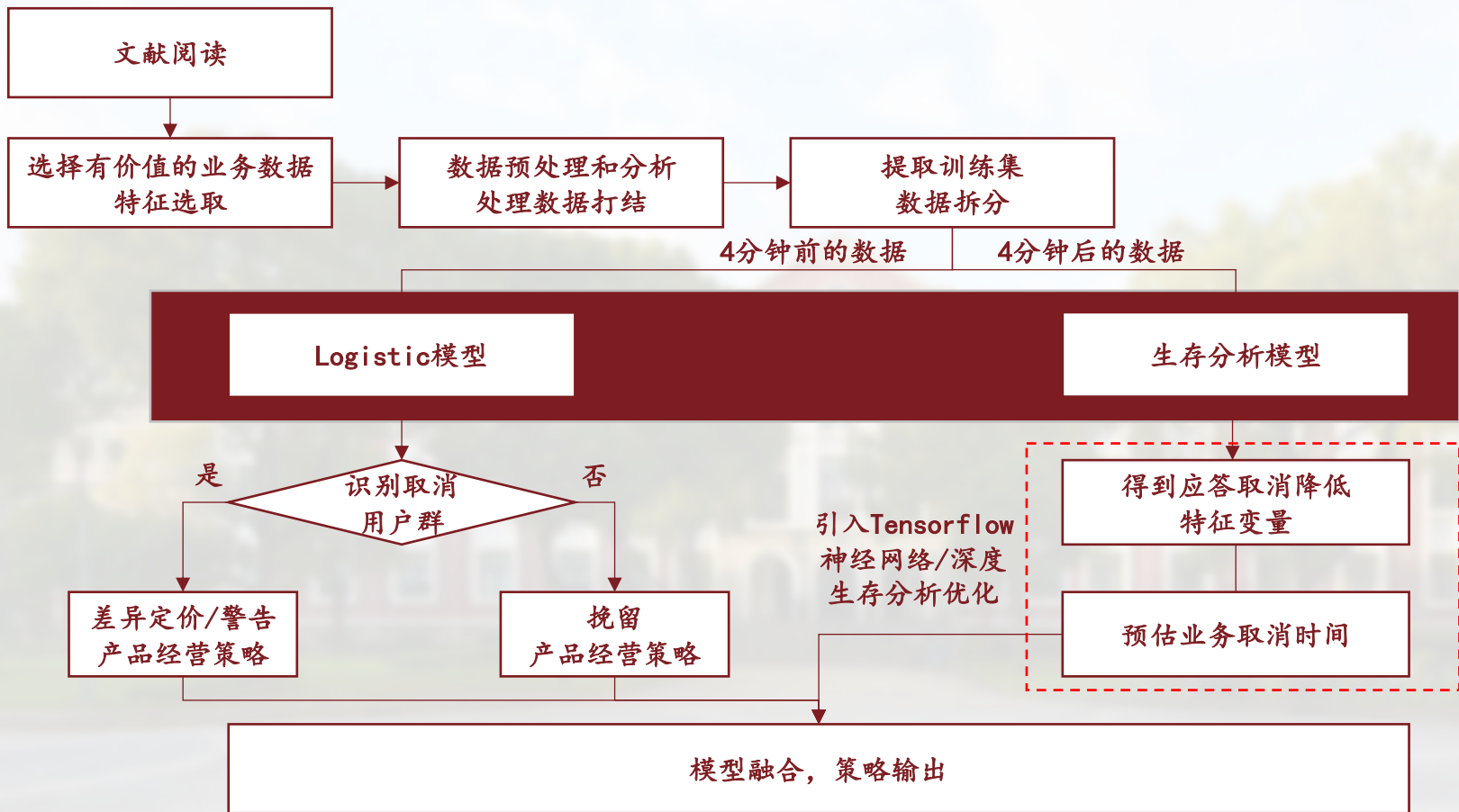
PART THREE

03

研究的基本思路及方法



上海财经大学
Shanghai University of Finance and Economics



Logistic回归

Logistic回归主要运用在流行病学，用于探寻某类疾病的危险因素。将因变量设置为二分类分线性差分方程，对于前4分钟的乘客应答取消问题较为适配，更加容易解释，也更容易帮助平台快速做出决策。同时还可以对不同自变量情况下的风险概率进行预测。

生存分析

生存分析很适合应用于乘客的应答取消研究，因为不仅考虑到终点时间的发生与否，同时也将终点事件出现的时间考虑在内，有效的解决了研究中时间因素所带来的影响。

传统生存分析是广泛应用在工程学、生态学、经济学领域的分析方法，如分析癌症患者存活周期、工程机械发生故障时间等，主要是研究协变量与事件发生时间之间的关系。其原理是，相较于其他传统的机器学习模型，有效利用的删失数据样本。

生存分析主要研究被观测对象的生存时间分布情况及其影响因素，生存时间的分布情况往往不能用生态分布进行描述，有时服从相关参数分布，有时却不服从任何分布。

相对风险模型属于半参数研究方法，模型包含参数、非参数两个部分，其中非参数部分指不限定基础危险率 (Baseline Hazard Function) 的形式，参数部分指建立解释变量、危险率两者关系的函数。

Cox比例风险模型（也称半参数模型）是相对风险模型的一个特例，1972年由英国统计学家Cox提出，是一种可以处理多个因素对生存时间影响的问题，也能处理删失数据的生存分析方法。**即在生存时间的分布未知的情况下，又要分析多个风险因素对生存时间的影响时，符合本次的数据条件。**

方法	假设条件	常用方法	适用性	结果
非参数模型	不需要进行参数分布假设，只需要对数据样本进行顺序排列，就可以估计被观测对象在不同生存时间下的生存率	乘积极限法	适用于数据样本量少或数据未分组的情况	得到时间点上生存函数、生存曲线呈现阶梯型
		寿命表法	适用于数据样本数量较大（如数以千计的量）或数据按区间分组的情况	
		特恩伯估计法	适用于时间的生存时间中含有完全数据、左右删失数据的情况	
参数模型	假设生存时间服从某种分布	指数分布、威布尔分布、广义F分布等	适用于已推测出生存时间满足何种分布的情况	生存函数和时间关系是平滑的下降曲线
半参数模型	只需要对生存函数部分参数进行假设	Cox比例风险模型	适用于生存时间的分布未知的情况下，又需要分析多个风险因素对生存时间的影响情况	兼有非参数模型和参数模型的特点

本次建模训练集样本为特定时间段的巴西圣保罗单城市的所有快车订单数据，数据量为487,0517，协变量特征提取分为两部分，线上订单特征与线下历史特征。

线上订单历史特征包括：



通过分析巴西市场订单的地理分布，我们可以看出，有61.12%的订单集中在前10个城市中，其中圣保罗占全国订单的19.97%，里约热内卢占全国订单的13.82%，是非常依赖单体城市发展的市场，这其中当然也包含巴西许多城市的自身原因，例如争议地区或者危险地区等等，使得城市呈现头部集中度高的特征。

特征名称	特征意义	处理规则	特征来源及状态
stat_hour_cut_0	晚高峰：17-22点		NA
stat_hour_cut_1	早高峰：6-11点	由左侧分箱规则分箱后经过one-hot处理	NA
stat_hour_cut_2	夜高峰：22-3点		NA
stat_hour_cut_3	平峰：其他时间		NA
source_type_1	正常派单	one-hot处理	rt, order表
source_type_2	重新指派		rt, order表
is_dynamic_0	没有动调	去掉异常值-1后one-hot处理	rt, order表
is_dynamic_1	动调		rt, order表
pre_total_amt	行程预估乘客所需支付金额		rt, order表
start_dest_dis	行程距离/m	去掉离群大异常值后归一化	rt, order表
forecast_dur	行程预估时间/min		rt, order表
start_dest_v ¹	预估行驶速度	由行程距离/预估时间后得到，去掉异常值后归一化	使用已有特征，用ether特征处理算子进行计算
amt_km ¹	预估每公里单价	由行程价格/行程距离后得到，去掉异常值后归一化	使用已有特征，用ether特征处理算子进行计算

1 为新计算指标，由提取特征计算后得到。

线下历史特征主要指乘客的历史行为特征，包括：

特征名称	特征意义	处理规则	特征来源及状态
days_since_reg	注册总天数	平均数取整后填补缺失值，归一化	写入RT或dufe
cnt_call_orders_total_fast	历史快车叫单数	平均数取整后填补缺失值，去掉离群大值，归一化	rt, order表
cnt_call_orders_7_d_fast	前七日内快车叫单数	平均数取整后填补缺失值，去掉离群大值，归一化	rt, order表
cnt_finish_orders_total_fast	历史快车完单数	平均数取整后填补缺失值，去掉离群大值，归一化	rt, order表
cnt_finish_orders_7_d_fast	前七日内快车完单数	平均数取整后填补缺失值，去掉离群大值，归一化	rt, order表
cnt_used_coupon_30_d_fast	前30日内使用优惠券数	平均数取整后填补缺失值，去掉离群大值，归一化	写入RT或dufe
cnt_used_coupon_7_d_fast	前7日内使用优惠券数	众数填补缺失值，去掉离群大值，归一化	写入RT或dufe
amt_used_coupon_7_d_fast	前7日内使用优惠券总金额	平均数取整后填补缺失值，去掉离群大值，归一化	写入RT或dufe
amt_used_coupon_30_d_fast	前30日内使用优惠券总金额	平均数取整后填补缺失值，去掉离群大值，归一化	写入RT或dufe
cnt_available_coupon	目前可用的优惠券数目	众数填补缺失值，去掉离群大值，归一化	写入RT或dufe
avg_amt_payable_finish_orders_30_d_fast	近30天完成订单单均应付	平均数取整后填补缺失值，去掉离群大值，归一化	rt, order表
total_fast_ratio ¹	历史完单率	历史完单数/历史叫单数（历史叫单数=0的话令其也为0）	使用已有特征，用ether特征处理算子进行计算
cnt_orders_ratio_7d ¹	七天内的完单率	七天完单数/七天叫单数（七天叫单数=0的话另其也为0）	使用已有特征，用ether特征处理算子进行计算

1 为新计算指标，由提取特征计算后得到。

将训练集样本与测试集样本进行同样的特征工程，包括异常值的处理，缺失值的填补，离散特征的one-hot，归一化的处理。

03 | 删失数据比例



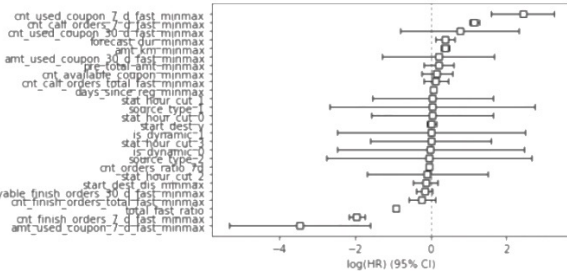
将训练集样本与测试集样本进行同样的特征工程，包括异常值的处理，缺失值的填补，离散特征的one-hot，归一化的处理。此时训练集与测试集的非删失数据比例为：

0	4387630	0	2809549
1	366493	1	244021

其中1代表非删失数据，0代表删失数据。

可以看到，非删失数据与删失数据比基本为12: 1左右。根据钱俊等人[1]针对删失数据比例对精确性的研究，在本次试验数据量下可认为此比例的生存分析结果依旧具有高可信性。最后将训练集经过欠采样使数据量降低为 (203603, 28)，非删失数据比例为：

0	142522
1	61081



	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%
stat_hour_cut_0	0.03	1.03	0.82	-1.58	1.64	0.21	5.15
stat_hour_cut_1	0.05	1.05	0.82	-1.56	1.65	0.21	5.23
stat_hour_cut_2	-0.09	0.91	0.82	-1.70	1.51	0.18	4.55
stat_hour_cut_3	0.00	1.00	0.82	-1.61	1.61	0.20	4.99
source_type_1	0.04	1.04	1.39	-2.68	2.75	0.07	15.69
source_type_2	-0.04	0.96	1.39	-2.75	2.68	0.06	14.59
is_dynamic_0	-0.01	0.99	1.27	-2.49	2.47	0.08	11.82
is_dynamic_1	0.01	1.01	1.27	-2.47	2.49	0.08	12.08
total_fast_ratio	-0.94	0.39	0.03	-0.99	-0.88	0.37	0.41
start_dest_v	0.02	1.02	0.06	-0.09	0.14	0.91	1.15
cnt_orders_ratio_7d	-0.05	0.96	0.01	-0.07	-0.02	0.93	0.98
pre_total_amt_minmax	0.20	1.22	0.20	-0.20	0.59	0.82	1.81
start_dest_dis_minmax	-0.14	0.87	0.17	-0.46	0.18	0.63	1.20
forecast_dur_minmax	0.38	1.47	0.13	0.13	0.64	1.14	1.89
days_since_reg_minmax	0.08	1.08	0.02	0.04	0.12	1.04	1.12
cnt_call_orders_total_fast_minmax	0.14	1.15	0.16	-0.18	0.46	0.83	1.58
cnt_call_orders_7_d_fast_minmax	1.16	3.18	0.07	1.02	1.29	2.79	3.63
cnt_finish_orders_total_fast_minmax	-0.23	0.79	0.18	-0.59	0.12	0.56	1.13
cnt_finish_orders_7_d_fast_minmax	-1.96	0.14	0.10	-2.17	-1.76	0.11	0.17
cnt_used_coupon_7_d_fast_minmax	2.43	11.40	0.43	1.60	3.27	4.95	26.25
cnt_used_coupon_30_d_fast_minmax	0.76	2.15	0.80	-0.81	2.33	0.45	10.33
amt_used_coupon_7_d_fast_minmax	-3.48	0.03	0.95	-5.35	-1.61	0.00	0.20
amt_used_coupon_30_d_fast_minmax	0.20	1.22	0.76	-1.29	1.68	0.27	5.38
cnt_available_coupon_minmax	0.16	1.18	0.21	-0.26	0.58	0.77	1.79
avg_amt_payable_finish_orders_30_d_fast_minmax	-0.17	0.84	0.11	-0.38	0.04	0.69	1.04
amt_km_minmax	0.38	1.46	0.06	0.27	0.49	1.31	1.63

Concordance 0.60

[1]钱俊. 生存分析中删失数据比例对Cox回归模型影响的研究[D]. 2009.

目前实验仍然存在一定的局限性，比如如上提到的训练集做完欠采样之后和原始数据集结果有偏差，研究颗粒度存在优化空间等。基于本文的一些研究试验和尝试，后续的研究工作方向主要有：

梯度下降 优化

由于深度生存分析损失函数的独特性质，目前只能做全批量的梯度下降，损失函数下降速度过慢，模型训练速度较慢。未来可考虑针对损失函数改进为小批量的梯度下降，增加模型训练速度；

特征多元化

当前特征为乘客线上订单特征以及乘客历史行为特征，可以考虑加入环境特征（如天气）、供需特征（考虑为时不变）等增加模型预测精确度；

增加训练集 规模

可增大训练集规模，目前的训练集为巴西某单一时间段的快车数据，在提升训练速度的基础上增加训练集样本量；

建立有效的 乘客流失挽 回方案

本文的研究仅针对准确挖掘大量特征的核心原因能力进行改进和研究，但在实际生意中，如何挽回这些取消乘客也非常重要。后续可以考虑通过阅读营销相关、心理学相关书籍文献，针对本文得到的使乘客应答前取消率降低的特征进行理解和学习，定制更有效的国际化乘客流失挽回方案。

PART FOUR

04

论文的创新及预期

CLICK HERE TO ADD TITLECLICK HERE TO ADD TITLECLICK
HERE TO ADD TITLECLICK HERE TO ADD TITLE



上海财经大学
Shanghai University of Finance and Economics

01 两阶段 问题研究方法

采用真实业务数据，根据实际以4分钟为节点的业务问题，对4分钟前、4分钟后的业务目标采用了差异化建模方法，4分钟前使用Logistic判别快速识别乘客取消风险，4分钟后采用生存分析同时判别乘客取消风险并记录生存时间，大大提高了司乘匹配效率、司机侧的运转效率，并有效节约了运营成本。

02 引入 Tensorflow 进行优化

由于网约车存在复杂的用户特征数据，而Cox回归模型仅利用的特征的线性信息，对非线性没有加以利用，因此我们可以通过在其中糅合深度学习方法^[1]，实现了非线性信息特征的利用。过往这种研究仅局限于医学领域。如果我们把乘客下单到取消作为一个时序特征数据，那么挖掘乘客的时序特征信息可以提高大数据集下的网络训练效率，同时也可以提高神经网络处理信息的能力。

[1] Liupei, TFDeepSurv, <https://github.com/liupei101/TFDeepSurv>

研究难点：

1) 数据打结问题处理。数据打结指的是多个被研究乘客具有相同的应答前取消时间。模型参数估计中，我们假设了某死亡时间没有重复事件发生，但是在实际应用中，尤其是乘客应答前取消在相同的时间内有大量事件发生，无法区分这些事件的时间先后。本文引入Exact Model, Discrete Model, Efron法以及Breslow法处理打结数据，最大化的利用了数据，增加了模型的置信度。

2) 遗漏重要变量。考虑到数据的加密和可获得性，本文的实证部分选取的协变量数量较少，研究结果可能遗漏重要变量。后续需要继续挖掘并将更多变量纳入考量范围，并基于特征筛选方法讲重要性高的变量加入建模过程。

PART FIVE

05

论文安排及目录

CLICK HERE TO ADD TITLECLICK HERE TO ADD TITLECLICK
HERE TO ADD TITLECLICK HERE TO ADD TITLE



上海财经大学
Shanghai University of Finance and Economics

- 先预计进行大量的文献阅读，对网约车相关问题实证研究、生存分析相关理论及模型问题、Logistic相关理论及模型问题、Pipeline建模方法进行系统性学习和梳理。
- 其次对数据进行探索性分析，对特征进行筛选、归类及预处理。多次尝试模型，对不同参模型进行比较（如Cox比例分析模型、深度学习生存分析模型）持续优化结果。
- 再次对目前产生的结果进行业务语言翻译和传导，阅读营销相关、心理学相关书籍文献，针对本文得到的使乘客应答前取消率降低的特征进行理解和学习，定制更有效的国际化乘客流失挽回方案。
- 最后预计输出模型结果、对应巴西市场乘客标签解决方案，并撰写论文。

- 第一部分为绪论，主要阐明选题的背景及理论意义和实用价值，并对国内外研究现状进行通盘梳理。
- 第二部分介绍采用的Logistic方法论及生存分析方法论，阐明使用的基本概念及定义，并对主要的模型方法如非参数法、参数法、半参数法等进行介绍和梳理。
- 第三部分为建模部分，首先阐明实际业务问题，选择合适的方法论，对数据进行处理，并对模型进行参数估计及特征训练。
- 第四部分为融合建模部分，针对前4分钟及后4分钟的建模结果进行融合，给到经营分析结论。
- 第五部分为结论及未来展望，对可以改进的地方进行盘点并估量下一步优化研究方向。

目录

第一章 引言	4
第一节 研究背景及意义	4
一、选题背景	4
二、理论意义及实用价值	4
第二节 国内外研究现状	4
第二章 生存分析	6
第一节 生存分析概述	6
第二节 生存分析相关定义	7
一、基本概念	7
二、生存时间	8
三、生存数据	8
第三节 主要生存分析模型	9
第三章 乘客应答前	11
第一节 生存分析建模	11
一、方法论选择	11
二、模型的参数估计	12
三、深度生存分析Cox模型	12
四、Cox回归模型对打结数据的处理	13
五、评价指标	14
第二节 模型的训练及验证过程	14
一、数据集及特征的选取	14
二、线性模型训练	16
三、深度生存分析模型训练	19
第三章 结论及未来展望	20
第一节 结论	20
第二节 未来展望及改进点	20
参考文献	21

目前研究进展



上海财经大学
Shanghai University of Finance and Economics



请各位老师们多多批评指正！
