

*使用 G.E.T 前请先阅读此文档

G.E.T Quick Model 说明文档

▼ 产品介绍

G.E.T 定位

G.E.T 机器智能要素

G.E.T 产品优势

▼ 快速入门

数据准备

模型训练

模型预测

▼ 系统配置

数据资源配置

离线任务管理

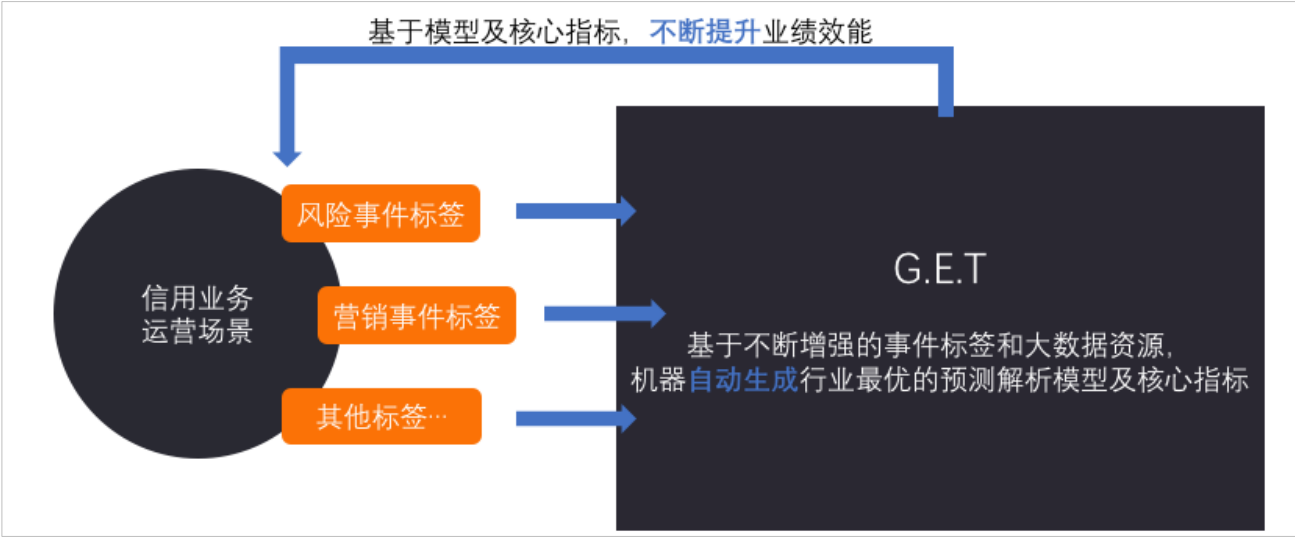
■ 可能遇见的问题

更新时间：2020.02

产品介绍

G.E.T 定位

G.E.T 定位为**微型智能建模机器**，能够基于不断增强的事件标签和大数据资源，机器自动生成行业最优的预测解析模型及核心指标，不断提升业绩效能。产品架构如下图：



G.E.T 机器智能要素

1. 无人力介入、不间断按照业务需求**自动调节**执行速度地处理数据（传统统计学与机器学习需要建模人员用丰富的经验来选择模型、筛选特征以及调节参数等，耗费大量时间避免过拟合与欠拟合）；
2. 能够识别**过挤**和**有空间提升**的数据信息内容；
3. 根据业务人员需求，灵活进行**交叉优化**，例如：控制风险，最大化意愿度；控制意愿度，最小化风险。

G.E.T 产品优势

1. 优良建模内核：领先于传统机器学习/评分卡的预测能力，快速完成**表现优良**的模型
2. 深度分析：提供 KS、AUC 以外的多类分析指标，**立体解读**人群、风险、利润等
3. 低门槛/高质量：面向算法人员的简易操作，无需资深金融建模人才，一样**获得高水准建模能力**
4. 高效生产：分布式计算，可视化训练/预测，**缩短建模周期 3 倍~10 倍**

快速入门

“数据准备、模型训练、模型预测” 三步完成 G.E.T 智能学习。

数据准备

请在**自有环境**完成训练样本准备（G.E.T 暂不提供开发平台），并将结果以 csv 或 txt 文件存储至本地。

训练样本数量 **30 万以内**，特征变量数量 **5 个以上、2000 个以内**，其他格式约定如下：

字段名称	字段类型	字段说明
id	string	主键（命名为 id，系统据此识别主键）
y_label	int	标签（命名为 y_label，系统据此识别事件标签）；“1” 为正例，“0” 为反例
feature1	int/float	特征变量（命名不限，此处“feature” 仅为示例）
feature2	int/float	
feature3	int/float	
.....	int/float	

注：字段无罗列顺序要求，G.E.T 程序仅根据字段名称辨别主键、标签、特征变量。

模型训练

在“训练”页面中间，自定义模型名称、提交训练集，再点击“GET 一下”执行训练：



点击“GET 一下”按钮后，系统自动弹出以下训练列表，待训练完成后，**刷新页面**显示 KS、AUC 核心指标；若需要详细考察训练成果，请点击“下载报告”，下载至本地阅读：

G.E.T Quick						
G.E.T训练记录						
项目ID	项目名称	创建人	创建时间	训练成果	预测管理	
78	test3	admin	2020-01-20 16:59:29	KS=0.50 AUC=0.81 下载报告	本地预测	离线预测
77	test3	admin	2020-01-20 16:50:14	KS=0.50 AUC=0.81 下载报告	本地预测	离线预测

模型训练分析报告的核心指标解读：

模块	文件路径	文件名	字段名	含义
训练	Otp\t\20200119105954\rslt_20200119105954\AICO_AFR_TRN	AICO_AFR_TRN_tdy_mdsg_20200119105954	auc	单变量的 auc
			grwth	对应 auc 估值的隐含变化率
			mean_rbst	鲁棒性平均估值
	Otp\t\20200119105954\rslt_20200119105954\AICO_AFR_TRN\pcmd_ms50_cr95_0	AICO_AFR_TRN_tdy_ai_sb_ac_perf_nb_pcmd_ms50_cr95_0_20200119105954	Sample_Size	训练集样本数量
			Bad_Size	训练集正样本数量
			AFR	训练集集正样本占比
			AUC	模型准确率
			KS	模型好坏区分程度
	Otp\t\20200119105954\rslt_20200119105954\AICO_AFR_TRN\3_co\nb	AICO_AFR_TRN_tdy_ai_sb_co_bgtb_nb_20200119105954	Total	模型输出概率排序后分段样本数量（限定 regime_no=1）
Bad			模型输出概率排序后分段正样本数量（限定 regime_no=1）	
Bin			模型输出概率排序后分段编号（限定 regime_no=1）	
测试	Otp\t\20200119105954\rslt_20200119105954	comd_ac_perf_20200119105954	Sample_Size	测试集样本数量
			Bad_Size	测试集正样本数量
			AFR	测试集正样本占比
			AUC	模型准确率
			KS	模型好坏区分程度
		comd_prd_02_20200119105954	pcmd_name_new	测试集 id 的预测值（可转化概率）

注：“文件路径”及“文件名”中出现的日期以实际日期为准。

模型预测

G.E.T 目前提供“本地预测”“离线预测”两类服务，未来将提供“在线预测”服务。

本地预测

本地预测适用于临时性、一次性的预测场景，对特定用户集快捷预测、分析预测结果；本地预测入口为：“预测”页面>预测管理>本地预测：



进入“本地预测”功能后，将特定用户集 的 id 及特征变量，以本地 CSV 或 TXT 文件提交，并执行预测：



本地预测数据集格式约定如下：

字段名称	字段类型	字段说明
id	string	主键（命名为 id，系统据此识别主键）
feature1	int/float	特征变量（命名不限，此处“feature”仅为示例）
feature2	int/float	
feature3	int/float	
.....	int/float	

注：字段无罗列顺序要求，G.E.T 程序仅根据字段名称辨别主键、特征变量；

离线预测

离线预测适用于批量性、例行事务性的预测场景，对大量用户（甚至全量用户）执行预测，且需要不断更新，以保障预测结果新鲜度；离线预测入口为：“预测”页面>预测管理>离线预测：



进入“离线预测”功能后，提交 SQL Query 脚本，G.E.T 将使用该脚本在数据仓库查询 id 及特征变量，执行预测，并将预测结果按指定库、表存储；

若无需更新预测结果，请将“更新频率”选为“从不”，否则按提示选择更新频率（成功提交部署后，系统将立即执行预测，此后按指定更新频率，每日凌晨离线执行）：



注：SQL Query 脚本请注意时间选择，若特征变量需要更新，可以将时间写为参数，读取服务器时间。

系统配置

数据资源配置

数据资源配置应用于创建 G.E.T 平台与数据资源的连接，入口如下：



“训练”和“本地预测”无需配置数据资源，离线预测需连接数据仓库，如下：

离线预测

当前支持hive文件(关系型数据库拓展中)。数据库连接配置：

数据库类型:

* 数据库地址:

* 端口:

用户名:

密码:

离线任务管理

离线任务管理应用于任务运行失败时便捷重启，入口如下：



面向任务状态为“失败”“中断”的任务，支持重启，如下：

离线预测_任务运行管理						
模型名称: <input type="text" value="名称模糊搜索"/> <input type="button" value="搜索"/>						
任务ID	模型名称	任务状态	计算进度	计算时长	提交时间	操作
35	test_train	失败	<div><div></div></div>	22s	2020-02-26 11:00:50	重新提交

可能遇到的问题

Q：模型训练、本地预测**执行失败**怎么办？

A：请检查提交的 CSV 或 TXT 文件的数据格式是否符合约定，若已排除该原因，请联系我司技术支持。

Q：离线预测**部署失败**怎么办？

A：请检查提交的 SQL 文件的脚本格式是否符合约定，若已排除该原因，有可能缘于集群计算资源紧张，可以稍后在“离线任务管理”功能中，重新提交部署，若仍无法解决异常，请联系我司技术支持。

Q：模型生命周期管理（**迭代**、**下线**等）怎么操作？

A：针对已经部署离线预测的模型，可以在“预测管理”功能模块可视化操作模型下线；当前 G.E.T 在 UI 界面未提供模型迭代的操作功能，若有需要迭代，请下载旧模型，新建新模型。