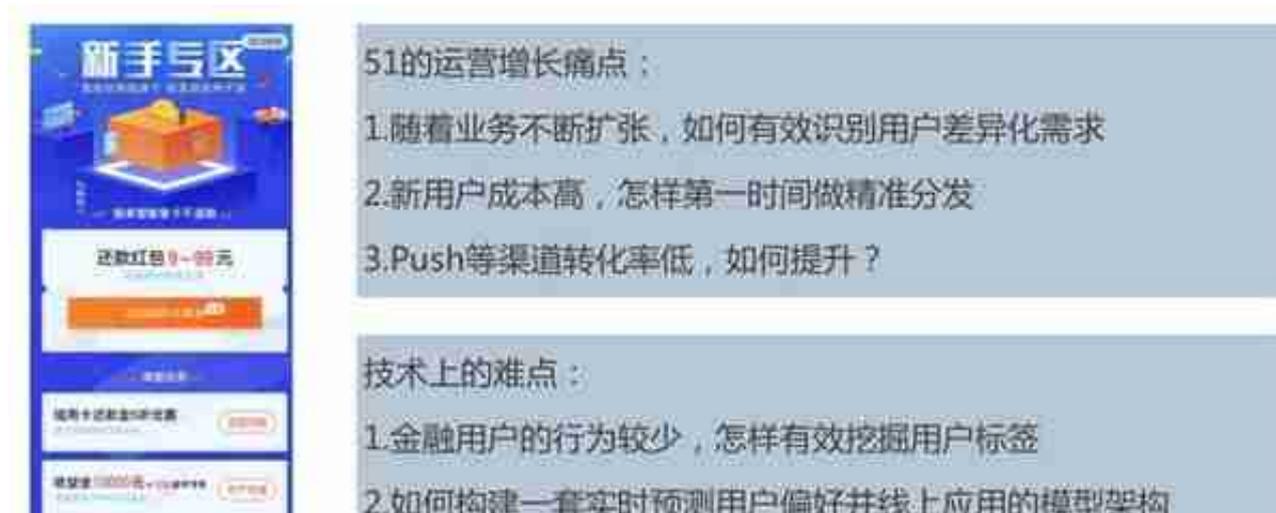


### 导读：

人工智能是当下金融产业转型的重要助力，本文主要介绍51信用卡管家在这方面如何搭建模型体系，如何发掘数据价值，如何成立自己私有的扫码平台或个性化体系。

51信用卡管家最开始以账单管理业务为主，国内比较早的用于管理信用卡账单的APP，目前公司的业务领域涉及负债管理、金融、技术服务等。



## 03

### 系统的架构平台相关内容

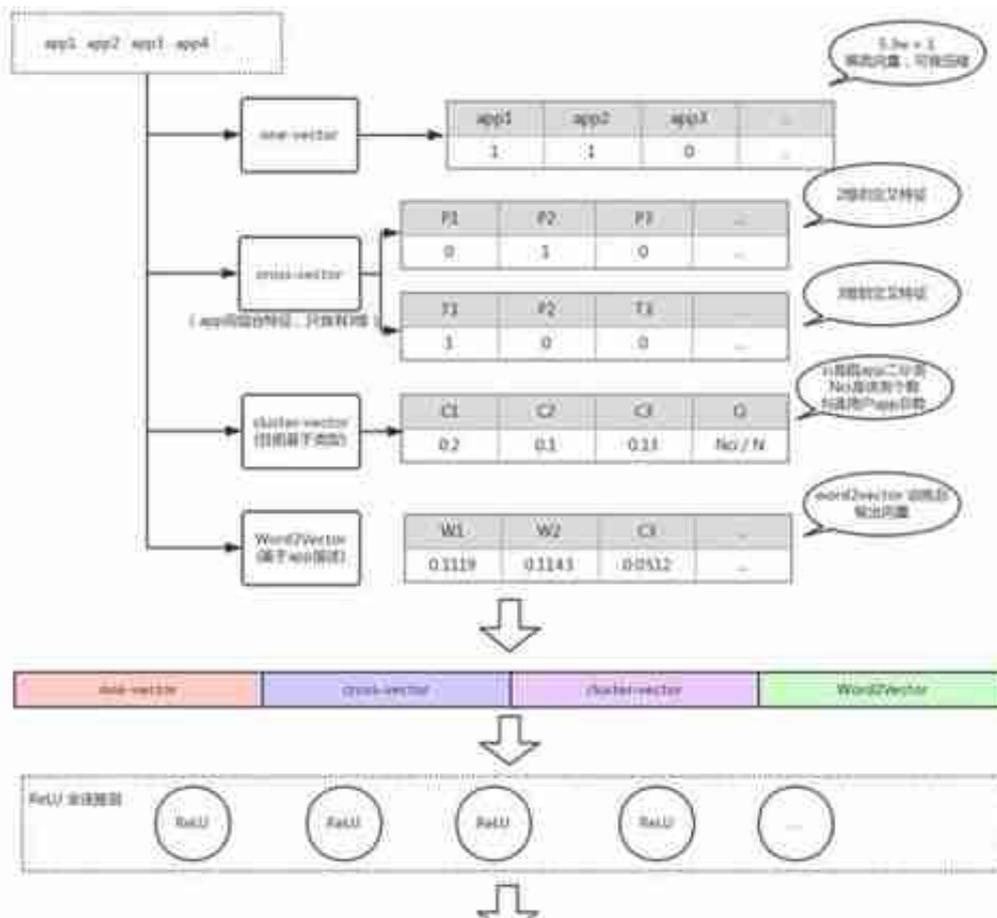
金融业务如何基于账单和APP构建用户标签？金融业务中数据来源第一个是账单，数据量较大，用户的很多账单需要管理和分析，提供很多定制化服务，比如推线下的很多活动等，这个对用户来说意义比较大，这里的账单主要是消费类的数据。第二类数据是APP数据，经过很多分析表明这类数据重要性比较高，在后面将结合具体算法介绍基于APP构建用户标签。

举例说明基于账单构建用户标签。假如今天在杨国福麻辣烫消费了一笔账单，账单里面可能会形成“支付宝消费—杨国福麻辣烫—\*\*金额”等没有语法结构的短文本，我们需要挖掘有用信息，经过很长时间的探索，我们发现这些短文本基本都有一个核心短语，将这些核心短语分类列出来对我们帮助非常大。

下图是我们设计的分词系统，大概有三层体系，第一个是主向量化，有字级别向量以及词级别向量化，词向量化也需要很多算法；再往上一层是合并规则，如果名词之间相关性比较大时合并起来就是短语，此外我们也有自有词库；再往下一层就是概率图推理模型，简单介绍一下这层的作用，对于账单来说有几大问题，假如有一笔套现账单，这种

账单在这情况下意义不大，但对于判断风险评估来说比较有用，比如判断用户偏好，这样的账单记录会影响系统分词效果。再比如有一笔信贷业务的账单，我们要判断用户是否具有还款或负债能力，此时具有欺诈性质的假账单影响比较大，需要筛选出来。总体原则是避免有问题账单进入系统而影响结果。

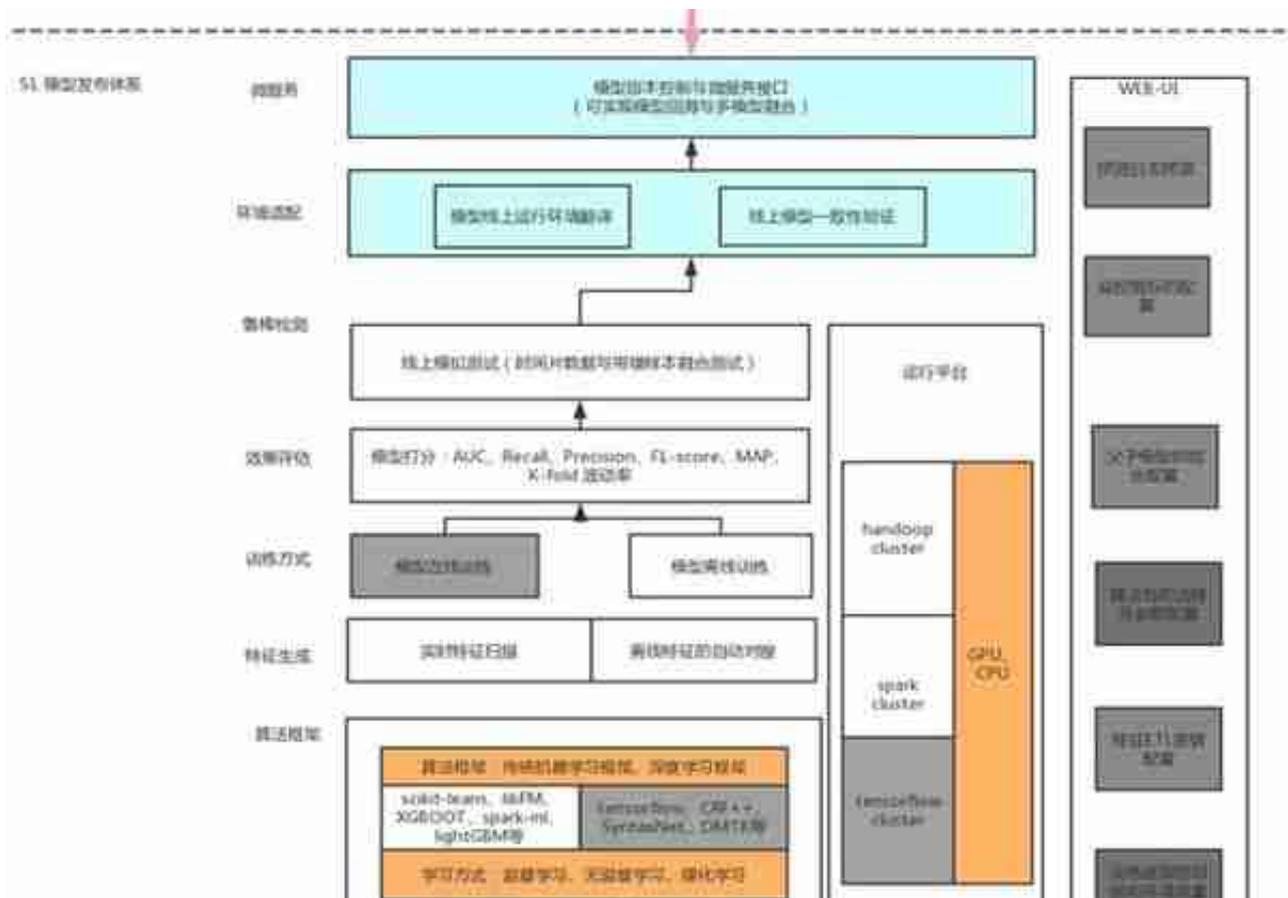
以“支付宝消费—杨国福麻辣烫—\*\*金额”为例继续介绍，如下图所示，先将账单记录分为两大类，关键词还是实体名，将“杨国福麻辣烫”提取出来，其它部分作为other部分。像“支付宝消费”这样的词在分词系统里面other部分一步步分离，最终进入系统概率图部分。概率图有什么作用呢？将训练过程历史记录中出现词的贡献概率或联合概率识别出来，即图中边与边之间的关系。假如进入这个概率图后形成四个短语，在四个图中各出现一次，其他图都没有出现的话，这种意味着这四个短语在图之间的关系在历史训练中比较弱，独立性比较强，那么我们看一下这四个图各处在哪里，与库里面其他词概率关系是怎么样的，然后构建模型。此外，我们也可以通过人工系统进行审核。下图右侧账单算法分词案例，首先将账单所有短语列出来，大的地区通过将中国的地理位置排列识别，小的地方通过其他解决办法也可识别。还有一个问题，商家都有明显的酒店短语，为何还要做呢？大多数账单能够识别出有多少是住宿类别，但有些时候识别不出怎么办呢，比如“香米拉客家商务酒店”，那么需要通过其它方式提供有效信息再进行分类。以上内容是基于账单体系架构构建用户标签的过程。



接下来介绍如何将这模型高效利用，各个业务线如何打通，建模人员如何高效利用起

来。

在金融领域这类介绍的比较少，接下来简单介绍一下我们的算法框架。我们算法上线时候，有一个统一的自动化翻译过程，不管是Python、C++还是R语言写的算法统统被屏蔽掉，因为我们要求线上环境必须是稳定的，此外产品上线情况下还需要做移植性校验，将翻译过程的错误进行封装。流程再往上走，51作为一个个性化通用平台，特征生成包括实时特征扫描、离线特征自动对接以及多元特征融合，训练方式包括模型在线训练、模型离线训练，效果评估包括模型打分指标：AUC、Recall、Precision、F1-score、MAP、K-fold波动率等，鲁棒性检测包括线上模拟测试(时间片数据与带噪样本融合测试)。再往上走就是环境适配，包括模型线上运行环境翻译与线上模型一致性验证。再往上就是51CG运营管理平台，实现了全自动化，因此我们模型上线速度特别快，只要配置好信息随时能够上线对接。

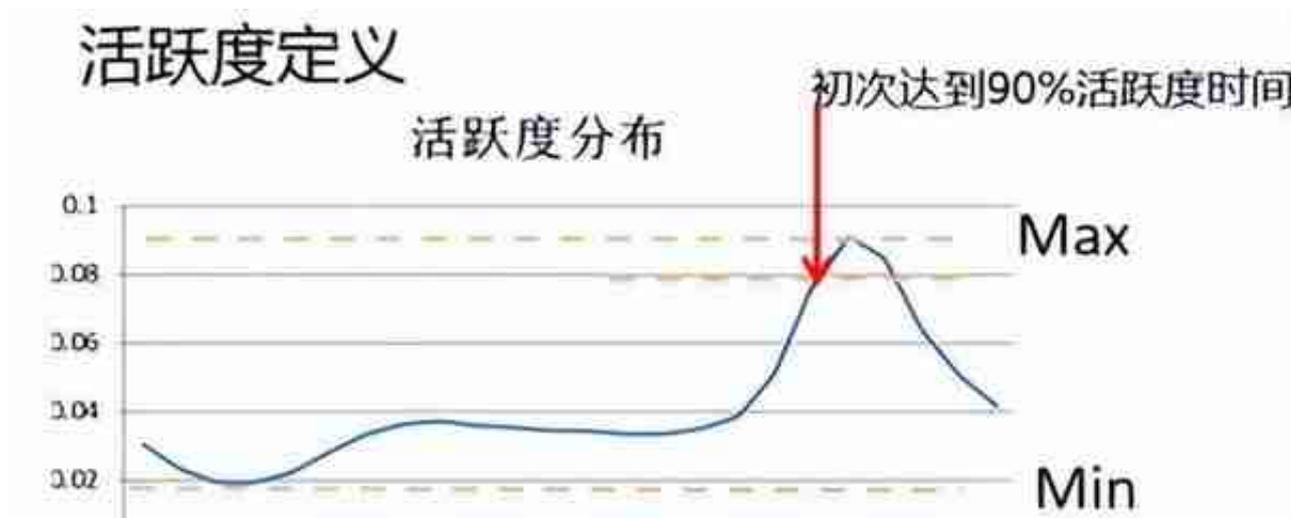


# 04

## 案例介绍

## 1. 增长的案例1

新手专区个性化AB。51管家最大的功能点就是管理账单，如下图所示针对多种类型新用户（导单、还款、理财、办卡贷款、征信）实现T+0（实时毫秒级）个性化新手专区AB测试新手专区转化率较默认业务排序提升7-14%。

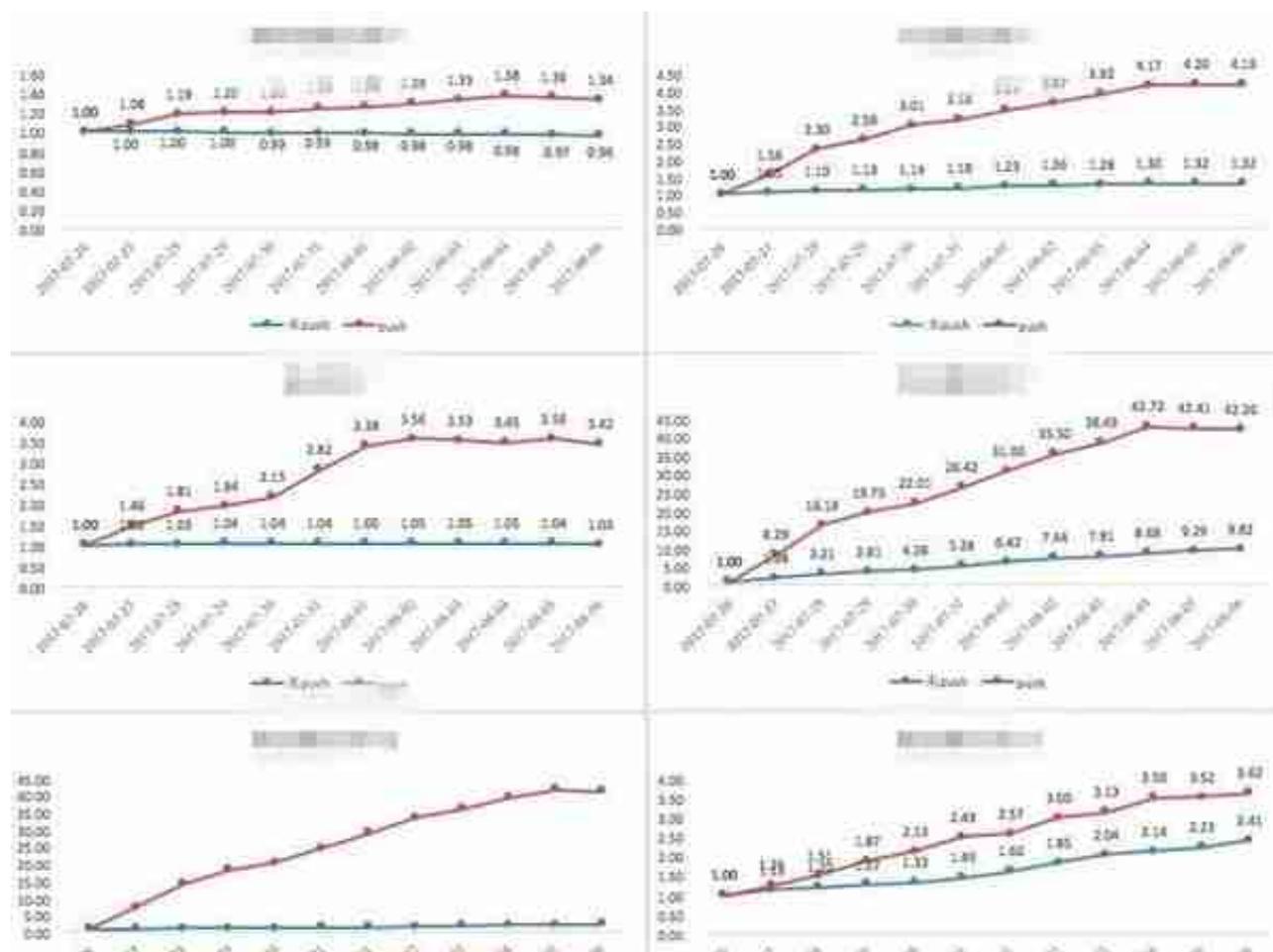


理财文案比较多，我们通过MAB算法来选择最佳文案，简单介绍一下我们如何应用的？MAB是一个算法模型，我们采用UCB算法（置信区间上界）来计算每一个文案下收益，置信区间上下限宽度需要注意一下，如果置信区间上下限很宽的话是存在风险的，由于实验不够可能每次都定位同一个位置，这个时候如何校验需要一定的拆分实验功底。

level	预估分档	疲劳度	打开率	占比	说明
1	0.7-1	每天一次	38.50%	0.50%	核心用户
2	0.6-0.7	2天一次	14.30%	1.50%	核心用户
3	0.5-0.6	4天一次	3.20%	9%	活跃用户
4	次新用户	4天一次	1.30%	5天内新人	次新用户
5	0.4-0.5	6天一次	0.49%	33%	次活用户

## 2. 增长案例2

基于用户标签对理财用户进行细分预测。我们对用户定义分群非常清楚，比如低活流失用户、高活流失用户、周期型用户、尝试型用户等，包括前面提到的Push在增加用户活跃量的条件下，用户打开APP时基本就是活跃的，那么这些是基于我们数据的积累。



今天的分享就到这里，谢谢大家。

阅读更多技术干货文章、下载讲师PPT，请关注[微信公众号“DataFunTalk”](#)。

分享嘉宾：陈兵强 51信用卡 资深算法专家

内容来源：DataFun AI Talk 《51信用卡的个性化推荐体系》

出品社区：DataFun

## 01/分享嘉宾

陈兵强, 51信用卡管家资深算法专家，曾担任美国nanjee公司算法研发总监、国内领先大数据公司的ID-Graph团队负责人，长期从事金融领域的ML、DL等相关算法研究和落地解决方案。

## 02/报名看直播 免费领PPT