



# 基于“客户画像”的支付机构 客户身份识别方法探析

文 \ 曹婕

上海汇付数据服务有限公司

随着支付机构的不断发展壮大，反洗钱工作已逐渐成为支付机构在业务开展过程中的风险合规管理重点。近年来，支付机构严格执行《中华人民共和国反洗钱法》、《金融机构反洗钱规定》、《非金融机构支付服务管理办法》以及《支付机构反洗钱和反恐怖融资管理办法》等相关法律法规的要求，员工的反洗钱意识不断增强，反洗钱工作能力不断提升，反洗钱工作机制不断完善，反洗钱工作相关工作体系以及平台系统均已完成初步的建设，但与传统金融机构相比，由于支付机构在客户群定位以及产品业务设计等方面存在较大差异，因此在开展相应客户身份识别工作时也较传统金融机构存在诸多困难，包括对客户身份识别措施的局限性，缺少身份证件验证渠道或途径（例如对个体户营业执照的核实）；持续性的客户身份识别难度较大（例如在进行可疑交易监测时，难以对网上自助注册客户开展后续调查工作）；此外，在与银行、商户合作时，还存在难以获得深层信息支持，要素不全等问题。

对客户身份进行识别的过程就像对客户进行画像一样，积累的各方面信息越多，相应的细节特征也就越明晰，进而对客户真实身份的判断也就愈加精准。目前，支付机构的主要客户群以小微企业居多，且多处于成长阶段，业务、运营管理体系不稳健，天生伴随着对外信息不完善、财报虚假、现金流不透明等高风险因素。一方面，市场的现状以及目标客户群的特征导致支付机构客户端所采集的信息匮乏且可信度不高。另一方面，由于非银金融市场本身的认证及征信体系不够健全，支付机构的业务、产品较传统金融行业创新速度较快且更为多样性，使得支付机构在信用风险、欺诈风险、洗钱风险、支付风险、账户风险等方面面临的风险维度与层级都较金融机构更为复杂。

基于此，本文提出一种能够在业务开展中帮助支付机构进行客户身份进行识别的方法——“客户画像”法。不同于传统基于用户主动提交材料来进行身份审查的方法，本文从客户以往的生活轨迹着手，旨在建立一套能够有效分析客户历史行为

习惯的“客户画像”规则，以准确解析客户真实的身份特征，探查其背后是否存在违法隐瞒，从而帮助支付机构更加准确清楚地了解自己的客户，规避后续的金融交易风险。

## 方法思路

对客户进行画像，即对客户的信息予以标签化，也就是经由支付机构通过内外多种渠道对给定客户的社会属性、生活习惯、行为习惯等主要数据进行汇总，并利用大数据技术进行分析和处理，进而勾勒出该客户相应的信息“轮廓”，以帮助支付机构快速和精准的识别客户的身份特征（如图1所示）。

整体来看，“客户画像”的完成可经由以下四个阶段：一是战略解读——即首先明确为客户画像的意义，例如为群体性的业务策略、风险政策等提供依据，以

此确定功能目标和效果预期。二是数据建模——即结合客户实际需求，找出相关的数据实体，以数据实体为中心规约数据维度类型和关联关系，形成符合客户实际情况的建模体系。三是维度分解——以客户、业务产品、支付场景、外部环境进行数据维度分解和列举，进而根据相关性原则，选取和战略目的相关的数据维度，以避免产生过多无用的数据对数据分析过程造成干扰。四是颗粒度划分——对于不同角色人员的需求，有针对性地设计各角色人员在客户画像工具中的使用功能和应用操作流程。

## 方法关键点

### 标签体系的构建

获取“客户画像”的重点在于能够对给定的客户赋予精准的“标签”。其中，每一个标签实质上都是经过高度抽象的、

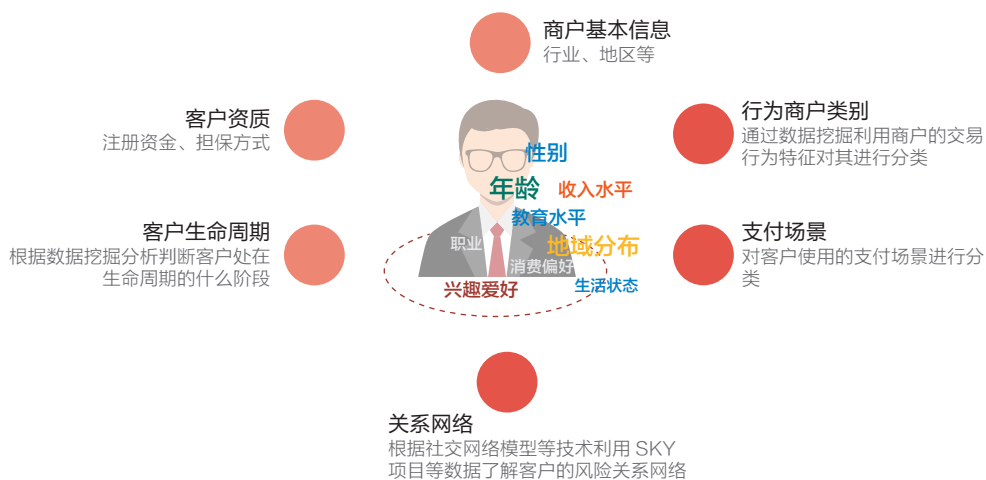


图1 “客户画像”实例



可量化的一个特征标识，如年龄/公司成立时间、个人收入/公司销售收入、性别、地域、客户行为偏好等。通过对给定客户获取并整合其所有的标签，即可完成对该客户“画像”的勾勒。

标签的获取来自支付机构对从外部或内部渠道获得的客户相关信息进行归类和分析所获得的一个或者多个具有代表性的字段信息。为了能够有效梳理标签，方便后续对数据深入挖掘，需要首先搭建一个完整清晰的标签体系。图2所示为一个典型的标签体系样例。其中，左边代表不同层级的标签数量，右边则是针对具体标签层级的解释说明。在该图的标签体系中，标签类别被划分为三个层级，其中，

一级标签是大类标签，如个人客户、机构客户；二级标签则是从属于一级标签的细分，如年龄/公司成立时间、个人收入/公司销售收入、性别、地域等固有属性；而三级标签是对二级标签的进一步细分，对应一级标签下的实体对象，如客户交易行为、消费习惯、所属环境因素等类。在标签体系确定之后，接下来即可据此对应不同层级完成客户群体数据的归类和细分。

### 数据挖掘

数据挖掘，即对客户标签体系中的数据进行深入分析挖掘，整个过程对应标签的层级采用由粗到精的处理方式，首先通过客户基础数据（客户的固有属性，如个

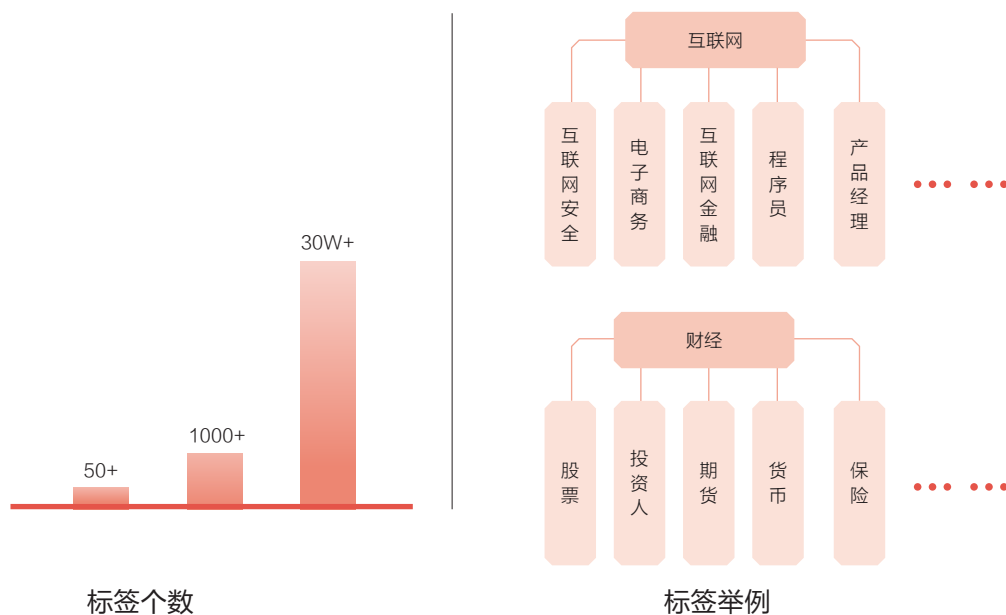


图2 三层客户标签体系样例

人收入/公司销售收入、性别等)，客户关系数据（不同客户之间的关联、共性特征等），客户行为数据(如客户的交易行为、消费行为等)，深入挖掘并赋予基础权重；其次通过验证信息作为权重的调权因子参与计算；接下来通过多个维度的定向挖掘系统和日常运营反馈系统进行校正和增加标签的覆盖；最后将挖掘出来的客户标签及权重输出至客户能力标签库，供上层业务调用（如图3所示）。

### 数据建模与应用场景聚焦

在获取客户能力标签之后，将结合具体的业务应用场景对客户的一步行为进行预判，对其中的潜在风险实施预警。首先针对具体应用场景的需求，依据不同

的评估维度，利用聚类算法（如K均值算法）将具有相近标签的客户进行聚类，对客户实施评级，以进一步精细化区分客户的特征。然后，对于同类客户进行建模分析（根据数据情况，模型可包含专家经验模型、混合模型与数据驱动模型），实施针对性的风险控制策略。

### 颗粒度划分

对于二维图像的应用来说，不同的图像分辨率对应于不同的应用场景。显然，对“客户画像”的颗粒度划分，即对客户行为特征的解析应该细化到何种程度，也应该与具体的业务应用场景相关联。这是因为，虽然对画像的颗粒度划分越细，被刻画的客户行为特征也就变得越发立体清

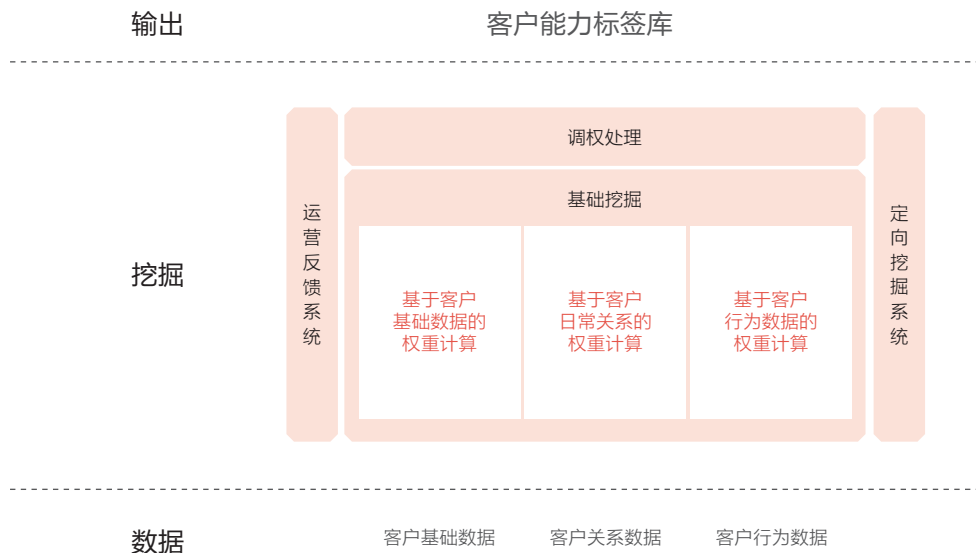


图3 客户标签数据挖掘框架



本文提出了一种基于“客户画像”的支付机构客户身份识别方法，该方法通过主动有效分析客户的历史行为习惯，避免了传统的依赖客户提交材料进行身份识别工作的弊端，从而帮助支付机构更清晰、更准确地了解自己的客户。

晰，理论上来说，该结果更有助于采用正确适当的风险管理措施。但是，过于细致的划分通常会导致所获取的客户信息琐碎化，弱化该客户原本具备的特定客户群共性特征，导致无法再使用大数据技术进行挖掘分析。同时，也因为过于细致的颗粒度，会造成目标过于单一（仅适用于该客户），出现过拟合（overfit）现象。此外，过度细致的颗粒度划分会带来大量的计算开销，导致运行成本攀升。

## 总结

本文提出了一种基于“客户画像”的支付机构客户身份识别方法。该方法通过主动有效分析客户的历史行为习惯，避免了传统的依赖客户提交材料进行身份识别工作的弊端，从而帮助支付机构更清晰、更准确地了解自己的客户。需要说明的是，本文主要是对所提出方法的原理思路和关键点进行论述，并未涉及具体模型和算法的设计，而这些将是我们下一步研究的重点。

### 参考文献：

- [1] 高国强. 非面对面客户的身份识别和风险控制 [J]. 现代经济信息, 2015, 09: 351 - 352.
- [2] 宋洪江, 隋作勇. 非面对面金融业务客户身份识别的难点及对策 [J]. 吉林金融研究, 2009, 09: 31 - 32.
- [3] 陈新英. 金融机构非面对面业务客户身份识别现状及对策 [J]. 金融科技时代, 2013, 11: 82 - 83.
- [4] 中国银行安徽省分行反洗钱工作调研小组. 对非面对面业务客户身份识别问题的思考 [J]. 决策, 2010, 01: 19 - 20.
- [5] 王禾立. 银行业金融机构非面对面业务反洗钱问题研究——以网银业务为例 [J]. 金融与经济, 2012, 11: 90 - 91.
- [6] 冯芸, 严畅, 杨冬梅, 张晶晶. 基于网络支付的电子商务环境中洗钱行为的识别和监管 [J]. 系统工程理论与实践, 2008, 12: 60 - 69.
- [7] 王勇. 探析“捆绑式”客户身份识别制度——基于第三方支付中的反洗钱视角 [J]. 浙江金融, 2009, 11: 13 - 14.
- [8] 张成虎, 岳鑫, 乐晖. 基于聚类方法的客户交易行为模式识别 [J]. 计算机工程与应用, 2007, 10: 195 - 198.