

# CFM AI&VR 报告：农商银行金融反欺诈项目交流报告

张义飞

UESTC — May 15, 2019

## 1 概要

2019年5月13日，本组成员张义飞与陈佳黎前往农商银行（白果林办公区）与银行方面张经理及其IT人员一名就“金融反欺诈”项目展开讨论。

## 2 背景信息

近期，由中国银行联合六部委发起一项行动，各地中小银行机构需向中国银行报备指定时间段内用户的大额交易数据，各地银行需对用户可疑交易行为进行监控并及时上报。为应对此行动，农商银行IT部门尝试引入机器学习方法用以实时监控用户交易流水，目前针对“拆分型地下钱庄”提出的线性分类器已取得十分良好的效果（对方声称准确率99%以上）。

在此基础上，农商银行方面希望继续使用机器学习方法对“商户偷税漏税”以及“结算型地下钱庄”等典型异常交易行为实现实时监控。

## 3 需求介绍：异常交易之反洗钱策略

### 3.1 案例：拆分型地下钱庄

**拆分型地下钱庄**是指洗钱者通过拆分交易来规避大额现金交易报告<sup>1</sup>，其主要特点是整存零取，通过高频小额交易来规避审查。目前农商银行的IT部门通过手动构造特征的方式，训练了对应的线性分类器，并通过集成学习等方法<sup>2</sup>，在线上<sup>3</sup>达到了99%的正确率。

### 3.2 合作内容

银行方面列举了三个可能的合作方向：**结算型地下钱庄、偷税漏税监控以及涉毒交易监控**。其中**涉毒交易监控**由于涉及公安部缉毒行动的保密信息，估计实行难度很大，所以对方希望在前两者中寻求合作。

关于**结算型地下钱庄**，对方没有进行太多介绍，根据笔者搜集的资料<sup>4</sup>得知，其交易模式大概为：洗钱者先是用很少或基本一致的资金注册多家公司（多为个体工商户），接着在一家或多家银行开立多个账户。其后通过该批账户的网银进行试探性交易，再迅速发展为超规模性交易。通过银行的网银、公转私、ATM转账和提现等功能业务，快速、大量转移资金，可以在短时间内将当地资金转移到外地，在外地通过数个过渡账户周转后再迅速回流回当地或指定地区取现，达到其最终目的。这类地下钱庄在前期开户时做足功课，使其开立的公司经营范围复杂，且涉及面广（如文具、服装、建材、五金、家电等），同时存在法人相同或关联法人、注册地址不存在或为家庭地址等特征。在银行开业账户时，选择开户地点集中、开户时间集中和代理开户人员相同或集中，且同时申请开通网银业务，以便统一控制和管理账户。非法结算型汇兑账户在运作资金过程中，公司账户表现为突然大进大出、不留余额；资金快进快出，日交易数十笔或上百笔；交易对手固定，不与其他单位或个人账户发生往来；虚假交易，交易对手与经营范围不符；分工明确，公司账户只办理转账，个人账户只接收汇款并通过各种途径提现。可以看到，结算型地下钱庄的洗钱模式同样有固定模式。

关于**偷税漏税监控**，对方列举了一个例子：生活中可见某些商店，明面上生意很差，几乎没有顾客前来购物，但却能持续运转很长时间，这样的商户往往是不法分子特地设立，用来做假账以逃避税务部门的监管。

经过初步讨论，对方希望初期针对偷税漏税监控与我们展开合作，后期再逐步扩展到其他方面。

<sup>1</sup>[http://www.sohu.com/a/127514283\\_465463](http://www.sohu.com/a/127514283_465463)

<sup>2</sup>对方并未提供过多细节

<sup>3</sup>根据对方透露的信息，可参考这篇新闻了解农商银行的大数据平台：[http://www.sohu.com/a/231836041\\_499199](http://www.sohu.com/a/231836041_499199)

<sup>4</sup>地下钱庄洗钱分析及防范建议：<https://www.xzbu.com/3/view-4543350.htm>

### 3.3 难点分析

与银行方面交流后，在监控偷税漏税方向，总结本项目难点在以下几个方面：

1. **标注数据过少**。对方透露，目前银行方面针对偷税漏税行为的标注数据仅仅涵盖十几笔交易数据，对方希望我方使用半监督的方法解决标注数据不足的问题。
2. **处理数据量大**。对方透露，目前已上线的拆分型地下钱庄行为监控日数据处理量是**数十万笔/天**，将来提出的机器学习算法将会部署在银行方的大数平台上。
3. **跨行交易信息缺失**。对方透露，用户的很多交易都以跨行的形式产生，比如跨行转账，而跨行转账流出的金额无法继续追踪，因为对方银行显然不会同意共享用户交易数据。用户流出的交易在他行多次周转后，很可能会回流到本行，但此时的交易特征肯定已经发生变化，所以对方希望我们能针对这一点能够提出创新性应对方法。

### 3.4 其他需求

对方IT人员提出了其他需求：

1. 对方想要了解，业界是否存在**数据自动升维**的方法，即**特则工程自动化**，用于省去手工构造特征的步骤。
2. 对方想要了解业界对于机器学习处理问题的最佳实践。

## 4 初步调研

笔者在沟通之后，针对对方提出的若干需求做了以下调研：

1. 针对**跨行交易信息缺失**的问题，可以考虑使用Graph Convolutional Network(GCN)对交易网络进行处理，但此方法可能存在计算量过大的缺点，可以作为一个创新点进行深入调研。
2. 针对**特征工程自动化**，业界与学界已经提出诸多方法，比如AutoML的子领域Neural Architecture Search(NAS)，以及业界已经存在相对成熟的特征工程自动化库FeatureTools<sup>5</sup>。

## 5 总结

在与银行方沟通时，对方表示希望与我们一同从创新性和实用性两个方面对机器学习在银行自动化监控方面的应用展开合作，其中创新性部分是指在对方提供的数据库基础上结合学界的最新方法进行创新，以产出论文的形式作为成果；实用性部分则指在线上环境取得良好效果。

---

<sup>5</sup><https://docs.featuretools.com/>