



云栖社区 (/) > AI早餐汇 (/teams/139) > 博客 (/teams/139/type\_blog) > 正文

://service.weibo.com/share/share.php?



(https://yq.aliyun.com/live/813)

宜人贷何林海：社交图谱在金融反欺诈产品的应用

技术小能手 (/users/30356932) 发布时间：2017-09-01 15:06:14 浏览2888 评论0

HTTPS (/tags/type\_blog-tagid\_456/)

电商 (/tags/type\_blog-tagid\_943/)

Image (/tags/type\_blog-tagid\_1019/)

数据采集 (/tags/type\_blog-tagid\_2061/)

知识图谱 (/tags/type\_blog-tagid\_17200/)

**摘要：** 在今年的「QingCloud Insight 2017」上，麦思博（msup）组织了主题为“AI落地实践”的专场论坛。宜人贷数据产品经理 何林海 在本场带来了题为《构建基于社交图谱关系的反欺诈产品应用》的演讲。

在今年的「QingCloud Insight 2017」上，麦思博（msup）组织了主题为“AI落地实践”的专场论坛。宜人贷数据产品经理 何林海 在本场带来了题为《构建基于社交图谱关系的反欺诈产品应用》的演讲。

何林海在演讲中主要分享了欺诈行业面临的挑战和产业链式发展、如何快速评估欺诈风险/平台搭建（框架体系）及实现、基于社交知识图谱的欺诈团伙发现、宜人贷面向行业的反欺诈解决方案“先知”等内容。

本文为早餐君根据速记整理而成，在不改变讲者原意的情况下做了编辑和缩略。特别致谢青云Qing Cloud、麦思博（msup）、宜信技术团队对本文的贡献。

何林海，目前主要负责宜人贷反欺诈产品的搭建，在做反欺诈之前，主要负责SDK行为数据的采集，收集了大量宝贵的用户行为数据，目前在公司内各个项目中得到了广泛使用。在宜人贷之前，就职于美团数据工程部。

以下为何林海的演讲内容。

我是来自于宜人贷的何林海，今天主要分享社交图谱关系在金融反欺诈产品方面的应用。

关于宜人贷

宜信公司2012年推出的一款在线金融服务平台。宜人贷通过科技驱动金融创新，为中国优质城市白领人群提供高效、便捷、个性化的信用结款咨询服务。

2015年12月18日，宜人贷在美国纽约证券交易所上市，成为中国金融科技第一股。截止2017年3月31日，宜人贷累计服务了超过60万借款用户，近120万出借用户，累计促成借款总额达393亿人民币。



(https://yq.aliyun.com/articles/680707)  
达人介绍



技术小能手 (/users/30356932)  
文章 7722篇 | 关注  
+ 关注

博主其他文章 更多> (/users/30356932)

- 一位技术校招生在支付宝的成长笔记 (/article)
- 阿里巴巴动物园又又又又来了新同学 (/article)
- 毕玄：我在阿里的十年技术感悟 (/articles/68679)
- 一枚戒指，一场仪式，这件事阿里巴巴坚持了 (/articles/688679)
- 2019年开发者必读！20位阿里技术大牛们帮你经典书单！ (/articles/688654)
- 智库大会 | 高红冰：拥抱智能时代的新市场经 (/articles/688566)
- 阿里云智能总裁张建锋：云更应该是CEO关 (/articles/688275)
- Euler 今日问世！国内首个工业级的图深度学 (/articles/688074)
- 天啦噜！在家和爱豆玩“剪刀石头布”，阿里工 (/articles/688052)
- 智库大会 | 智能科技：从万物互联到万物智能 (/articles/686945)

相关话题 更多> (/articles/688052)

- 又双叻叻删库了，这次是 AI 动的手 (/roundta
- 啥是佩奇，为何刷屏？ (/roundtable/493870)
- 性能测试 PTS 有奖分享，新年大礼包等你拿 (/roundtable/493729)
- 那些年你用过哪些消息中间件？为什么又弃了 (/roundtable/493090)
- 每天都用CDN，它的日志数据有这么大价值， (/roundtable/488717)

在美国，欺诈主要是由于信用风险而产生，但在国内，很多坏账是由欺诈产生，中国对于欺诈的惩罚不够，导致了类似事件频繁出现。作为一个行业的挑战，欺诈不仅只存在于在线P2P领域。

关于欺诈的行业挑战，举三个例子。

非银信贷：行业整体风险偏高，由于门槛较低，滋生了大量团伙欺诈、身份冒用、老赖、多平台借款、黑中介等，带来巨大经济损失，坏账率居高不下。

银行：开展网络支付、直销理财、电商、消费信贷、P2P等创新业务中，和非银信贷面临同样的问题。

电商：电商领域的在线支付、恶意点击、广告监测、黄牛、账户安全等方面面临交易欺诈风险。

这些问题是全行业的挑战，宜人贷面临的主要是信贷申请时候的欺诈。

现在大量欺诈已经不再只是个人作为，而是一种行业性质的行为。我简单分成三类：上游、中游、下游。



上游，主要是黑产技术服务，能力不逊于技术团队，力量强悍，积累了丰富的工具使用经验，包括虚拟模拟资料，做帐单造假等，成员都是计算机高手。

中游，把恶意注册的帐号释放出去，或者是通过交流平台招揽技术人才。另外，由于需要很多资料，到农村收一些大爷大妈的身份证。对他们来说，赚一百块钱，身份证借用两天没什么关系，实际上这些资料被用来骗贷了。

下游，有专人用这些资料来做欺诈、盗窃、勒索，甚至用来刷单。

### 如何快速评估欺诈风险及平台搭建

宜人贷如何建立一套快速评估欺诈风险的机制呢？

2014年开始有这些想法，在现有风控体系外，把欺诈风险和信用风险分开做。欺诈风险单独拎出来，并将对用户的欺诈风险识别出来。我们建立了一套从数据采集到机器学习，并快速给出欺诈评分的体系。



宜人贷何林海：社交图谱在金融反欺诈产品中的应用

### 如何快速评估欺诈风险？

**数据采集**：通过SDK采集设备信息、行为数据、交易数据等，在关键点采集用户行为事件信息。发送的数据越多，能够更加清晰判断该事件的欺诈风险大小，如激活APP事件、登录事件、注册事件、进行事件等。

**机器学习**：在海量事件、通过机器学习模型的训练，在更加精准地标注的模型，通过自动化地优化算法，通过监督学习、深度学习等多种模型的组合使用，有着非常好的稳定性与准确性。

**欺诈评分**：对每一个事件得出一个分值在300-900之间的欺诈评分，评分越高，其欺诈风险越低。客户可通过欺诈评分以及给出的标记对用户进行判断，同时为客户提供了自动化的决策引擎，对于高风险用户，系统会自动拦截，对于低风险用户，系统会自动放行。

(<https://yq.aliyun.com/articles/680707>)

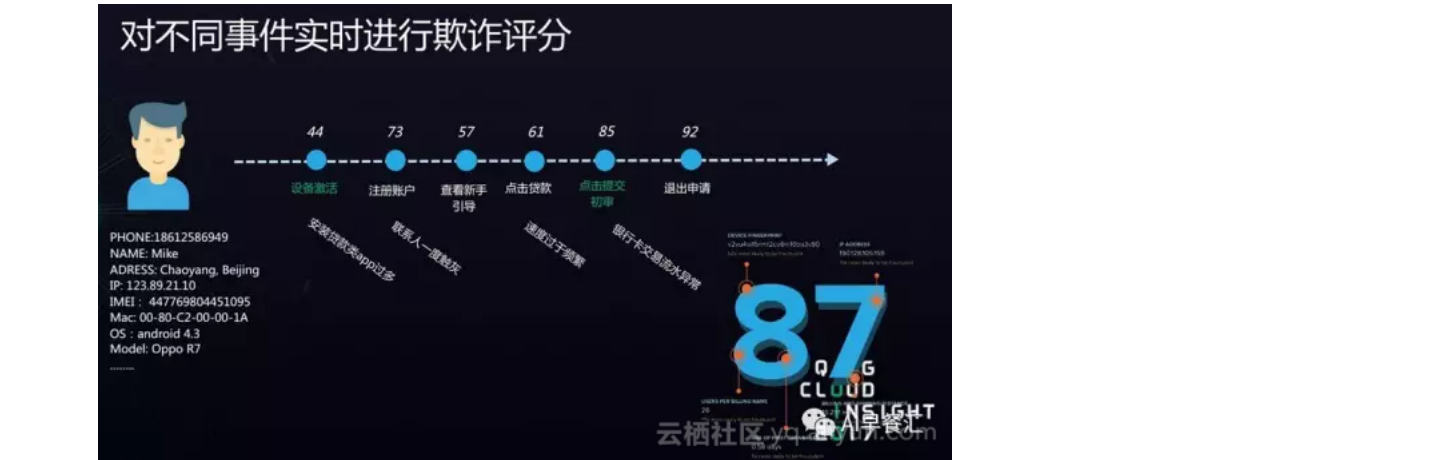
://se...om/...php?...

%E5...E4%BA%BA%E8%...BD%95%E6%9E%97%E6%

在数据采集时，通过SDK的方式，也就是采集SDK，可以在不同触点采集到用户的基本信息，只要用户触达任何一个终端，我们都实时地将这些用户信息送到SENT EVENTS上。

深度学习以前没怎么用，监督类学习是基于欺诈和非欺诈的用户去做分类，无监督用了知识图谱的技术。对于单个用户，每个事件会给出一个评分。

比如，用户打开APP会得分，因为怀疑安装了模拟器，用户进行申请时会给他提醒完善信息。对于优秀用户，根据他的设备或者关联出的其他信息，在注册时给他打一个好的分数。

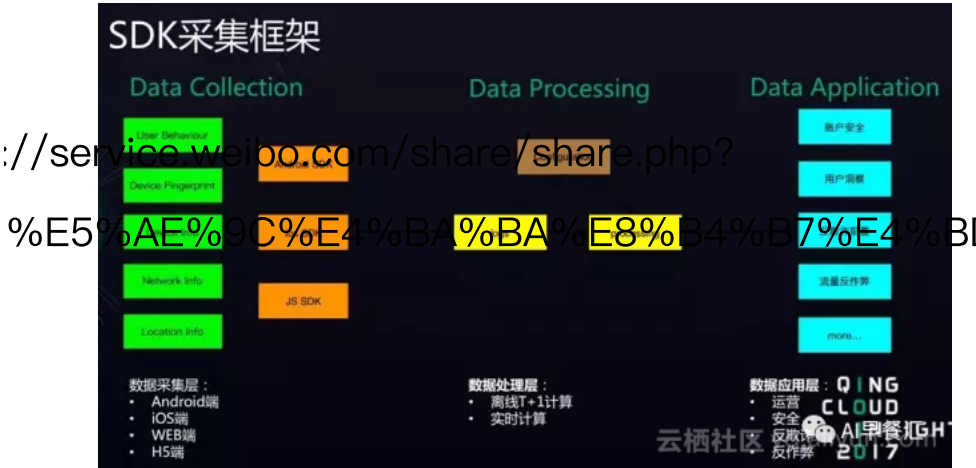


这是对于不同事件进行实时欺诈评分的流程。用户在设备激活的时候得一个分，注册的时候得一个分，查看新手引导的时候又得到一个分，这个分会实时变化。每次过来，用户都会带上基本信息：设备、IP地址等等，可用作评估风险程度的特征。



上面说到的是反欺诈的基本架构，我们在内部做成一个产品化平台，今天主要介绍一下它的体系，**宜人贷何林海：社交图谱在金融反欺诈产品中的应用**主要基于三个层次。

首先是实时的数据采集，数据采集分几个部分，第一部分是基于SDK采集相关的用户行为数据，第二部分是用户授权的相应个人信息，包括消费数据和通话数据，第三部分是三方数据。



其次，我们用到了实时数据处理，基于知识图谱构建了三个层次，第一个是欺诈评分模型，实时对用户的每个事件进行评分，第二个是传统的规则引擎，我们用了一个非常强的规则；第三个是团伙挖掘，这是一个很大的亮点。

前两点很多公司都在做。针对单个分类的问题，数据决定了模型好坏的上限，模型调的再好，只是决定效果无限逼近上限。团伙挖掘是基于知识图谱，把用户通过图的方式，挖掘他们之间的关系，把这些（有类似欺诈行为或关系的）用户拎出来。

在实时欺诈发现阶段我们采用产品化的工具，在流程里通过调查工具，直接给到信审。把每个用户提报出来，通过这套流程可视化平台，帮助快速理解用户是否欺诈，我们会做中介发现，把中介找到。第三是团伙监控和预警，把欺诈团伙挖出来。

### 每个阶段的具体实现过程

下面讲讲每个部分的具体实现过程。

第一个模块是基于SDK的实时数据采集，获取一手数据。我在去年加入了宜人贷，当时就在想，这么多的用户行为数据为什么不用来做反欺诈呢？用户的淘宝数据、消费数据，以及其他的很多临时性数据都可以作假，但是用户的行为数据一般不会去模仿，我们做了一个SDK采集框架，分成三层。

第一部分是数据采集，覆盖宜人贷所有客户触达终端的采集能力，采集用户行为、设备指纹等相关信息，比如型号、是否模拟器、各项物理层，以及网络源信息，这部分非常重要，它连接着一些上网轨迹，还有Location，我们发现很多团伙都在一个地点，好几十个坏用户聚在一个基站下面。以上是采集层。

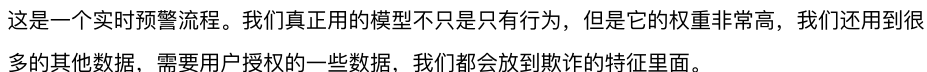
中间是实时数据处理，我们做了一层相应的配置，对于哪些数据是可用的，我们花了很多力气做数据清洗，刚开始的时候在这方面踩了不少坑，然后后续实时的时候会做一些数据回填，才达到一个比较好的运算。数据处理层分为离线和实时。实时数据直接灌到知识图谱里，离线数据做一些特征处理和建模分析。



采集的内容分为几个模块。设备数据主要有手机品牌、手机型号、操作系统、设备ID、App列表。

位置数据现在主要用四个：GPS、IP、基站信息、WIFI列表。比较精准的是GPS和WIFI，基站大概有100~200米误差。很多欺诈分子喜欢用代理IP。

基于以上，把用户实时行为数据采集完之后，对每个用户会形成一个实时的欺诈评分。



我们做了一个FICO评分标准，评分越高用户信用风险就会越小。此外还有一个规则引擎，我们会发现一些强有力的单条规则，或者多条单一规则组成的策略，对每个预警用户非常有效。

预警出来会分成两部分，一部分信用好的就直接拒了，另一部分交给反欺诈调研，形成黑名单。所有流程都是实时的，保证跟业务系统完全解耦，有快的反馈速度，十分钟放款，就要判断出他是否是欺诈。

模型训练，我们有一个好的点给大家分享一下，大家可能做过，正负样本非常不平衡，欺诈领域也有这个问题，我们面临的是想要找出一些坏用户，但很不易。

环用户常见的定义就是MOB6里面的M3+（逾期三个月），不能把有信用风险的用户剔除掉，这个用户不一定是欺诈，他是违约了，但有可能就是没钱还，要把欺诈跟信用解耦出，可以信用风险的方式做信用定价。我们构建了基于欺诈标注的自动化学学习，可疑对实时新增的欺诈标注，快速反馈回模型训练。

## 基于知识图谱的欺诈团伙发现



以上是对识别单个欺诈用户的应用，接下来谈谈如何基于知识图谱判断欺诈团伙，大家知道团伙其实很难发现，发现了也很难认定。

### 搭建基于Neo4j的图谱关系网

- Neo4j – Neo Technology, Inc.
  - 完整的ACID支持 – 原生的图存储和处理
    - 原子性 (Atomicity, 或称不可分割性)、一致性 (Consistency)、隔离性 (Isolation, 又称独立性)、持久性 (Durability)
  - 高可用性、有集群模式
  - 轻易扩展到上亿级别的节点和关系
    - V2.3.3 的节点上限是340亿
    - V3.x.x 没有节点上限, 只限于硬盘和内存
  - 通过便利工具高速检索数据
    - Java, .NET, JavaScript, Python, Ruby

云栖社区 yq.aliyun.com



neo4j



云栖社区  
CODING LIFE

(<https://yq.aliyun.com/articles/680707>)

我们采用了两种方式，第一种是通过无监督的方式，通过聚类的方式找到这些用户；另外就是通过知识图谱的方式发现关联关系。我们的知识图谱关系网用的是Neo4j。

我们主要应用知识图谱做了三件事情：疑似中介识别、关联骗贷团伙识别、团伙监控&预警。关联骗贷团伙，一个用户来了之后，对单个用户的特征评价里有了非常多的数据，他跟某一个团的关联关系，他在团里的关联特征，比如一度联系人，和二度的设备联系有多少，这些是作为单个用户而言。此外还有对所有用户的全局搜索，从而发现骗贷团伙。

关于团的定义可以参考下图。

### 如何定义一个团

- 种子：从某一个结点出发发现一个团，该结点称为种子
- 骨干点：某一个团中的核心结点，是团的连接中心
- 无辜点：相对孤立的点，几乎与其他点没有连接
- 超级点：super node，该点的连接边数巨大
- 核介系数：团内边数除以  $(n * (n-1) / 2)$ ，n：结点数
- 聚集系数：描述一个结点的所有邻居结点的核介系数
- 团合并：两个团有一定比例的重合，可合并
- 连接强度：边的权重
- 连接频度：边的条数

云栖社区 yq.aliyun.com

以下是上线之后的短期效果。

### 短期效果

56%  
欺诈认定率

数千万  
已挽回潜在欺诈损失

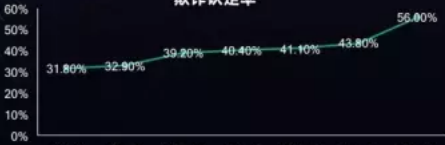
反欺诈模型

反欺诈规则

图谱事件提报

调查工具

欺诈认定率



Month	Fraud Identification Rate
JAN-17	31.80%
FEB-17	32.90%
MAR-17	39.20%
APR-17	40.40%
MAY-17	41.10%
JUN-17	43.80%
JUL-17	56.00%

- 采集SDK数据用于识别欺诈事件
- 使用机器学习反欺诈建模技术融合多种模型
- 结合人工调查与机器学习主动发现欺诈标注
- 构建知识图谱应用于中介识别和团伙识别

云栖社区 yq.aliyun.com

宜人贷何林海：社交图谱在金融反欺诈产品中的应用  
面向行业的反欺诈解决方案：“先知”

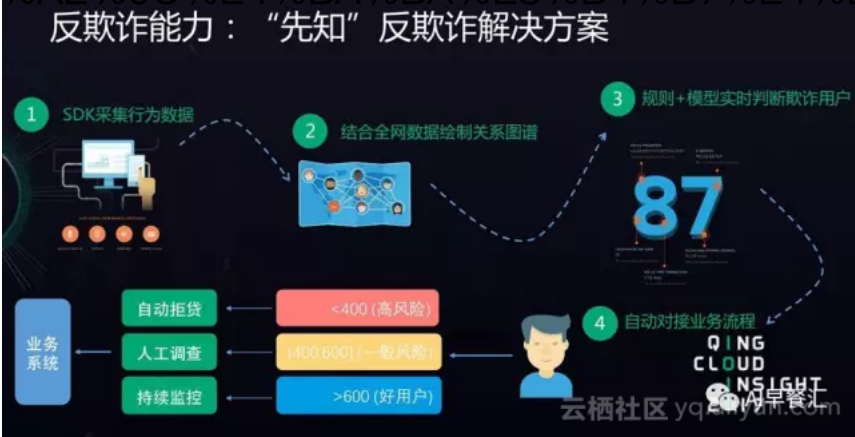
宜人贷面向全行业的一个反欺诈解决方案——先知。

宜人贷CEO对此提出了三项服务要求。第一是数据抓取服务，包括对于淘宝数据、公积金、社保数据抓取的服务，提供给其他行业；第二是反欺诈能力。第三是精准获客能力。

宜人贷有非常多存量用户，我们已经投了很多钱去获取基本用户，这部分用户客单价非常高，我们希望把一些不符合宜人贷定位，同时又是优质的客户，他的风险不高，但是他可能就想借一千块钱的用户，这种用户我们是转出去给其他的平台。



反欺诈能力：“先知”反欺诈解决方案



我们的反欺诈能力是通用的解决方案。第一，基于SDK采集行为数据；第二，结合全网数据绘制关系图谱；第三，通过规则+模型实时判断欺诈用户；第四，自动对接业务流程。

在产品流程方面，是基于SaaS的服务，用户可以在平台上自动完成预警欺诈用户设定，以及对于预警用户的处理。同时，这些服务可以对接到自己的业务流程里去，比如接入了“先知”系统后，对于用户预警，可以在平台上操作，同时也可以把这些数据通过API的形式对接到自有的信审里去。

原文发布时间为：2017-08-31

本文作者：Jack

本文来自云栖社区合作伙伴“AI早餐汇”，了解相关信息可以关注“AI早餐汇”微信公众号

如果您发现本社区中有涉嫌抄袭的内容，欢迎发送邮件至：yqgroup@service.aliyun.com 进行举报，并提供相关证据，一经查实，本社区将立刻删除涉嫌侵权内容。

【云栖快讯】云栖专辑 | 阿里开发者们的20个感悟，一通百通 详情请点击 (https://yq.aliyun.com/articles/680707)

评论 (0) 点赞 (0) 收藏 (0) 分享到: 微博 微信

上一篇: 高红冰: 1个贫困县连接280个城市 电商减贫潜力巨大 (... 下一篇: 用数据找到撬动新零售的支点 (/articles/192849)

(http://service.weibo.com/share/share.php?url=https://yq.aliyun.com/articles/188140&title=%E5%AE%9C%E4%BA%


相关文章

- 氮信资深数据科学家主讲：如何构建基于AI的金融风控系统 ... (/... 关于深度学习，这里有一份入门公开课（文末福利） (/articles/5...
- 宜信大数据负责人Joyce：当金融遇上大数据 (/articles/78773) 宜信大数据负责人Joyce：当金融遇上大数据，能擦出什么... (/a...
- 包银消费CTO汤向军：消费金融大数据风控架构与实践 (/articles/... 互联网金融，如何用知识图谱识别欺诈行为 (/articles/81179)
- 中国人工智能学会通讯——AI如何深度应用于消费金融的风控..... 互联网金融2.0：数据生态与战略制高点 (/articles/178593)
- 深度|普惠大数据中心首席数据科学家李文哲：知识图谱的应用 (/... OSDI '18重磅解密：蚂蚁金服实时金融级分布式图数据... (/articl...

网友评论

宜人贷何林海：社交图谱在金融反欺诈产品中的应用

登录后可评论，请 [登录](https://account.aliyun.com/login/login.htm?)



<https://yq.aliyun.com/articles/680707>

<://service.weibo.com/share/share.php?>

- 精华导读

限时流量包 (<https://promotion.aliyun.com/next/q/honorbaund.htm>) 云存储 (<https://www.aliyun.com/product/oss>) 网络安全 (<https://www.aliyun.com/product/aliyuncs>) 云数据库 (<https://www.aliyun.com/product/rds>) ECS升级配置 (<https://yq.aliyun.com/ask/33742>) 物联网 (<https://www.aliyun.com/product/iot>) 教程 (<https://edu.aliyun.com/taocheng>) PHP (<https://yq.aliyun.com/php>)
- 用户关注

自动化测试 (<https://bbs.aliyun.com/read/301499.html>) 解决方案 (<https://www.aliyun.com/solution/all>) linux命令 (<https://yq.aliyun.com/articles/34777>) 云服务 (<https://www.aliyun.com/>) JavaScript 函数 (<https://yq.aliyun.com/articles/92145>) 服务器监控 (<https://yq.aliyun.com/articles/48786>) Python语言 (<https://yq.aliyun.com/roundtable/56407>) 移动数据分析 (<https://www.aliyun.com/product/man>)
- 更多推荐

用户体验 (<https://yq.aliyun.com/articles/132294>) 云数据库Rds (<https://help.aliyun.com/product/26090.html>) 负载均衡 (<https://www.aliyun.com/product/slb/>) 域名注册 (<https://wanwang.aliyun.com/whois>) Whois查询 (<https://whois.aliyun.com>) 数据可视化 (<https://help.aliyun.com/product/43570.html>) CP备案查询 (<https://beian.aliyun.com>) 主题地图 (<https://yq.aliyun.com/zt>) 阿里云大学 (<https://e-cn域名>) (<https://wanwang.aliyun.com/domain/cn/>) Js (<https://yq.aliyun.com/jsarticle>) Mysql (<https://yq.aliyun.com/sqlarticle>) 移动站 (<https://m.aliyun.com/yunqi/>) IT论坛 (<https://bbs.aliyun.com>) 企业邮箱 (<https://mail.aliyun.com/>)

关于我们 (<http://www.aliyun.com/about>)      法律声明及隐私权政策 ([https://terms.aliyun.com/legal-agreement/terms/suit\\_bu1\\_ali\\_cloud/suit\\_bu1\\_ali\\_cloud201710161525\\_98396.html](https://terms.aliyun.com/legal-agreement/terms/suit_bu1_ali_cloud/suit_bu1_ali_cloud201710161525_98396.html))      廉正举报 (<https://jubao.alibaba.com/index.html?site=ALIYUN>) (<http://www.aliyun.com/contact>)      加入阿里云 (<http://www.aliyun.com/careers>)

阿里巴巴集团 (<http://www.alibabagroup.com/cn/global/home>)   淘宝网 (<http://www.taobao.com/>)   天猫 (<http://www.tmall.com/>)   聚划算 (<http://ju.taobao.com/>)   全球速卖通 (<http://www.aliexp.com>)   阿里巴巴国际交易市场 (<http://www.alibaba.com/>)   1688 (<http://www.1688.com/>)   阿里妈妈 (<http://www.alimama.com/index.htm>)   飞猪 (<http://www.fliggy.com>)   阿里云计算 (<http://www.aliyun.com>)   AliOS (<http://www.alios.cn/>)   阿里通信 (<http://aliqin.tmall.com/>)   万网 (<http://wanwang.aliyun.com/>)   高德 (<http://www.autonavi.com/>)   UC (<http://www.uc.cn/>)   友盟 (<http://www.umeng.com/>)   虾米 (<http://www.xiami.com/>)   优酷 (<http://www.youku.com/>)   钉钉 (<http://www.dingtalk.com/?lfrom=20150205111943449>)   支付宝 (<https://www.alipay.com/>)   达摩院 (<https://damo.alil.com>)

© 2009-2019 Aliyun.com 版权所有 ICP证：浙B2-20080101

 (<http://dinfo.zjaic.gov.cn/bscx.do?method=lzxx&id=3301963301080000025024>)

 浙公网安备 33010002000099号(<http://www.beian.gov.cn/portal/registerSystemInfo?recordcode=33010002000099>)