# 使用TigerGraph提高机器学习的准确性

## 机器学习的崛起

根据国际数据公司（IDC）预测，到 2021 年，在人工智能和机器学习上的投资将从 2017 年的 120 亿美元增长到 576 亿美元（数据来源：福布斯杂志，2018 年 2 月）。机器学习正在被应用于各种领域，例如反欺诈，反洗钱（AML）以及在线销售中的产品推荐。然而，想让当前的机器学习来识别一些诸如欺诈或洗钱的异常行为犹如大海捞针；因为为了找到大海中的“针头”（例如欺诈者），我们必须对海量的数据进行分类和解析。假设有一家电话公司，它的网络中每周产生数十亿次的呼叫，我们应该如何设计算法，让机器可以学会在海量的数据中找到欺诈行为的线索呢？

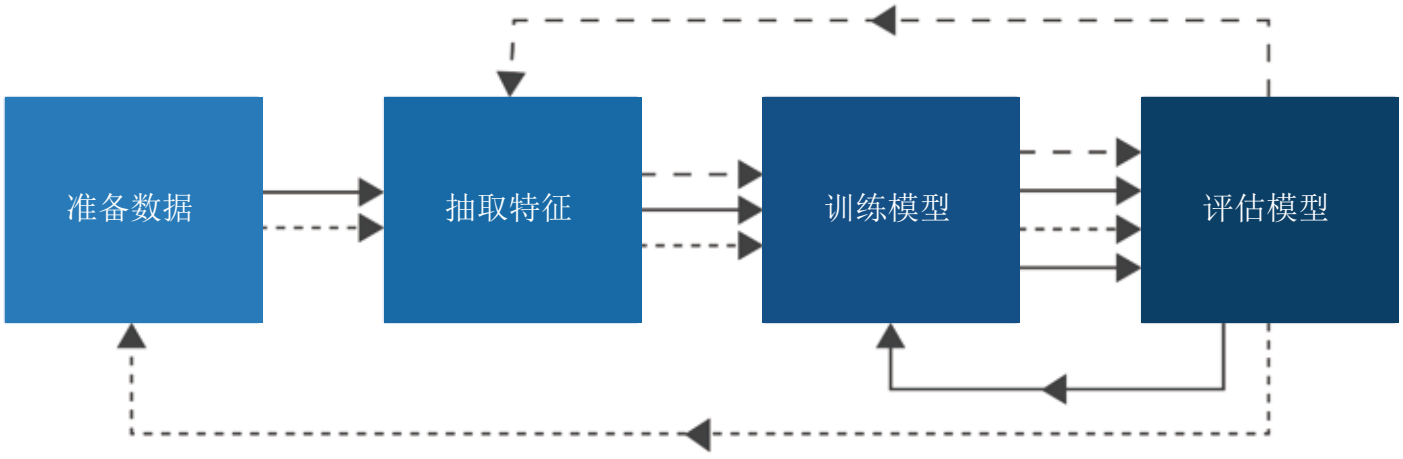
## 现有训练机器学习的方法缺陷

让我们通过这个电话公司的例子，来探究一下当下通过机器学习识别欺诈者的方法。现有的机器学习算法依靠于原始训练数据——例如在上述案例中，训练数据就是被已经确证的欺诈电话。但是还需面对两个问题——训练数据的数量和质量。

由于现有号码网络中能被确证的欺诈号码量不足总呼叫量的 1%，因此可以作为训练数据的欺诈电话量也屈指可数，这从而降低了机器学习算法的准确性。

当下的反欺诈功能是基于对某些行为特征或属性的简单分析。这些特征包括了，某个号码与其他网内外号码的通话记录，某张预付费 SIM 卡的使用时长，单向呼叫的百分比（指被呼叫方未回电的情况）和被拒绝接听的百分比等。这些过于简单的特征分析往往会导致大量误报，因为除了欺诈者之外，销售人员或恶作剧者也常常会有类似行为特征。

机器学习交互流程



## 基于图特性训练机器学习，进行欺诈侦测

某大型移动运营商使用 TigerGraph 提供的，具备实时深度链接分析功能（Real-Time Deep Link Analytics）的新一代图数据库来弥补当前机器学习算法的缺陷，将该方案运用在了包括 4.6 亿部手机在内的 100 多亿次呼叫分析之中，并为每部手机生成 118 个新的图形特征。这些特征基于对通话记录的深入分析，跨越了电话的直接接听者而直接延展到整个通话网络。

下图示意说明了图数据库如何将手机号码识别为“可信号码”或“嫌疑号码”。被定义为“嫌疑号码”的记录会被进一步调查以确证它是否的确属于属于欺诈电话。

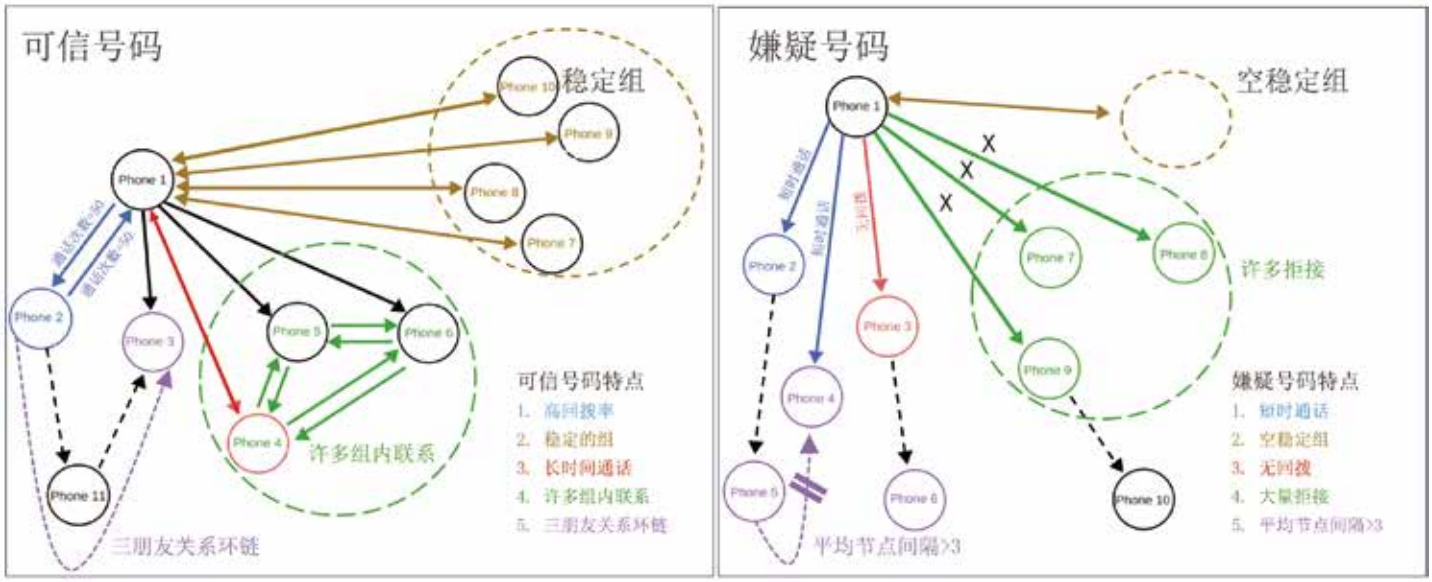



图 1 – 通过分析网络或图形关系特征来检测电话欺诈行为

“可信号码”与“嫌疑号码”

“可信号码”的第一个特征是，多数都会有对方回电，这显示出号码用户之间的熟悉或信任关系。一部“可信号码”也会每周或每月定期拨打一系列其他的号码，并且在一段时间内保持稳定（称之为“稳定组”）。

“可信号码”的另一个特征是，某个号码呼叫另一个已经入网达多月或多年的号码并得到回电。若某些“可信号码”、长期入网号码以及一些与它们有频繁联系的号码之间有大量的通话记录，这也是个很好的信号，它代表“可信号码”有着相当多的组内关联（in-group connection）。

最后的特征是，“可信号码”往往符合“三朋友关系环链”（three step friend connection）—— 即当号码 1 呼叫号码 2，号码 2 呼叫号码 3 时，号码 1 与号码 3 也有直接的通话。因为这种关系环链隐含着一个相互信任和关联的圈子。

通过分析这些号码之间的通话模式，TigerGraph 可轻松识别那些可能涉及欺诈的“嫌疑号码”。这些嫌疑号码总是会在短时间拨打多个可信号码，却不会被回拨。同时它们也没有定期稳定的通话组（即“空稳定组”）。嫌疑号码也不会被长期入网的号码客户回拨，同时时常被拒接，也缺乏三朋友关系环链。

总结一下这个案例：TigerGraph 对每部电话创建了超过 118 项特征属性，通过对 4.6 亿部电话相关联属性的分析，将这些电话区分为可信号码或嫌疑号码。与此同时，它新产生的 540 亿条数据，可以作为训练数据为机器学习算法的自我提升提供支持。这使得通过机器学习进行欺诈检测的准确性大幅提高，并同时降低了误报率（将非欺诈号码标记为欺诈号码）和漏报率（涉及欺诈的号码未被标记）。

## 基于图的特性提高机器学习的准确性

让我们通过一个示例（图 2）来了解一下图的特性是如何提高机器学习的准确性的：假设有四个移动电话用户的数据，他们分别是蒂姆，莎拉，弗雷德和约翰。

基于图的特性提高机器学习的准确性


| | | | | |
|---|---|---|---|---|
| SIM卡使用时间 | 2周 | 4周 | 3周 | 2周 |
| 单向通话比例 | 50% | 10% | 55% | 60% |
| 被拒接比例 | 40% | 5% | 28% | 25% |
| **机器学习根据历史记录预测的结果** | 疑似欺诈者 | 普通用户 | 疑似欺诈者 | 疑似欺诈者 |
| 稳定组 | 是 | 是 | 否 | 否 |
| 许多组内关联 | 否 | 是 | 否 | 是 |
| 三朋友关系环链 | 否 | 是 | 否 | 是 |
| **机器学习根据深度链路分析得到的分析结果** | 疑似恶作剧者 | 普通用户 | 疑似欺诈者 | 疑似销售人员 |

按照传统的通话记录属性（如 SIM 卡的使用时长，单向通话的百分比以及被拒接百分比），会导致他们四人中的三人被标记为疑似欺诈者。从这些传统属性来看，蒂姆，弗雷德和约翰都非常像一个潜在的欺诈者。但是基于图特性而做的深层链接或多跳关系分析，则可以帮助机器学习将他们区分开来，识别出蒂姆是恶作剧者，约翰是销售人员，而只有弗雷德才会被标记为潜在的欺诈者。

因为蒂姆拥有一个“稳定组”，这意味着他不太可能是一名销售人员，因为销售人员每周都会拨打不同的电话号码。而蒂姆又没有很多组内关联，这意味着他很可能经常给陌生人打电话。同时他也没有任何三朋友关系环链来证明他与他的拨打对象彼此认识。因此根据这些判断，蒂姆很可能是一个恶作剧者。

约翰没有一个“稳定组”，这意味着他每天都在拨打陌生人的号码。但同时他却拥有许多组内关联。当约翰通过电话向他的客户推荐产品或服务时，如果他的客户认为产品或服务对他们有价值，那么他们中的某些就会将这些产品服务推荐给他们的朋友。这样约翰就因此建立起了三朋友关系环链。这表明约翰作为一名优秀的销售人员，可以通过一轮针对朋友或者同事的销售，将产品或服务信息传递出去，形成一个关系链闭环。这些特性使得约翰能够被识别为一名销售人员。

反之，弗雷德即没有一个稳定组，也没有与任何拥有组内关联的群体有过互动。同时弗雷德与他的拨叫用户之间也没有三朋友关系环链。这使得他最可能成为电信诈骗的调查对象。

回到之前那个海底捞针的比喻——通过基于图特性的分析，我们便可以在海量的数据中找到那根针。在上述的案例中，弗莱德便是那个潜在的欺诈者，也就是我们要找的针。通过使用图数据库，我们可以对相互关联的数据进行分析并识别某些特征，而同时机器本身也可以得到大量高相关性的训练数据（基于图形特征的数据），使之在识别潜在欺诈者方面变得更加智能和成功。

# 使用TigerGraph提高机器学习的准确性

## 了解更多：

### 基于图特性训练机器学习的一些其它适用案例

除了用于识别电话诈骗之外，TigerGraph 实时生成的图特性也同时被用于大量其它场景，这其中包括训练机器学习算法以检测各种其他类型的异常行为——包括所有在线零售商家都面对的销售产品或服务时的信用卡诈骗行为，以及横跨了整个金融服务生态系统的洗钱行为。这些违法行为的影响面涉及到银行，支付服务提供商以及当下新兴的数字货币市场（例如比特币，以太坊以及瑞波币（Ripple）等）。

### 开启属于您自身的更加智能的机器学习系统

线上零售企业也可以利用图的特性来分析客户的购买行为，从而更精准地把产品推荐给客户、客户的朋友以及其他拥有类似行为特征的人群。同时，新生成的图特性也可作为现有机器学习的训练数据，使得未来的产品推荐更加精准。


### 联系我们：

www.tigergraph.com.cn

sales@tigergraph.com


## 客户与应用案例

UBER VISA Alipay elementum STATE GRID CORPORATION OF CHINA wish

TigerGraph 的实时大图分析引擎帮助世界最大的移动支付企业进行后台防欺诈分析，帮助世界最大的电子商务企业生成产品推荐，以及帮助世界最大的电网公司进行网络管理。

### 反欺诈和反洗钱

TigerGraph 的深度链路分析和大图分析能力能够挖掘出过去难以发现模式与关联。反金融犯罪部门可以通过实时的图形建模，调查可疑交易，高风险顾客以及相关关联关系。

### 客户智能服务

TigerGraph 帮助企业迅速实施高性能的客户关系分析系统。实时分析能力使得在线零售商可以迅速综合与感知客户行为，智能分类产品并向客户做出实时化的个性商品推荐。

### 大规模交易处理

某世界最大之一的电子支付公司使用 TigerGraph 处理超过 1000 亿顶点的图和每天超过 20 亿次的实时交易更新。该系统已在包含 20 个节点的生产系统中运行超过两年，达到完整的 ACID 规范。

### 智能电网

通过与业界领先的电力企业合作，TigerGraph 通过其革命性的本地并行图功能，帮助企业监控和分析电流，找到网络瓶颈并在网络出现性能问题时，及时向对应人员告警。

### 供应链智能管理

TigerGraph 提供针对关键供应链管理的实时可视分析工具，业务包含订单管理，送货管理以及其他物流业务

## 关于 TigerGraph

TigerGraph 特有的原生并行图（NPG）技术使其成为世界最快的图分析平台。面对实时处理极度复杂数据的挑战，TigerGraph 的图分析平台依然能够无视数据的海量性与复杂性，为客户达成预期并创造价值。TigerGraph 支持的应用包括物联网，人工智能以及机器学习，它们都通过感知瞬息万变的数据来实现。TigerGraph 的成功案例包括了支付宝、软银集团、中国国家电网、Wish 电商平台以及 Elementum 公司。您可以关注 TigerGraph 的微信账号 TigerGraph 或访问我们的网站 www.tigergraph.com.cn 获取更多信息。