你担心的金融安全 无监督机器学习技术可以搞定

人工智能3 月 20 日

日前，世界科技开发者盛会 DeveloperWeek 2019 评选 VR、人工智能、金融科技等领域优胜者，AI 公司 DataVisor 维择科技凭借无监督机器学习技术获得最具投资价值的科技金融企业奖。

“无科技，不金融”。随着移动互联网时代的到来，科技金融模式不断创新，但是欺诈手法也在不断翻新，呈现出专业化、产业化、隐蔽化等特点。日前，世界科技开发者盛会 DeveloperWeek 2019 评选 VR、人工智能、金融科技等领域优胜者，AI 公司 DataVisor 维择科技凭借无监督机器学习技术获得最具投资价值的科技金融企业奖。

无监督机器学习技术是什么，为何会被认为最具投资价值？它能在科技金融活动中起到什么作用？能解决哪些金融交易中的问题？

科技金融反欺诈创新利器

与传统金融不同，互联网金融业务大多发生在线上，往往几秒钟就完成审核、申请、放款等，面临的欺诈风险也是前所未有的。据统计，我国网络犯罪导致的损失占GDP0．63%，一年损失金额高达 4000 多亿人民币。国际上的情况也不乐观，多份市场研究报告指出，仅 2016 年一年，全球信用卡、借记卡、预付卡和私有品牌支付卡损失就高达 163 .1 亿美元；每年保险欺诈(不包括健康险)损失总额预计超过 400 亿美元。

“随着技术不断演进，针对金融业的攻击、欺诈手段已不同以往。团伙作案、分工明确、掌握各种先进技术工具、不断变化攻击手段，全新挑战使得金融企业越来越难以招架。” DataVisor 中国区总经理吴中说，金融反欺诈期待创新已成业内共识。

“无监督机器学习是近年才发展起来的反欺诈手法。目前国内反欺诈金融服务主要是应用黑白名单、有监督学习和无监督机器学习的方法来实现。”爱信诺征信有限公司总经理金端峰在接受科技日报记者专访时说。

黑白名单被认为是最原始的反欺诈方式，类似于“筛选器”。如银行征信系统就可理解成一个黑白名单，信用卡多次逾期还款就可能被列入信贷“黑名单”；在淘宝上购买了退货险后屡屡退货，就可能上骗保“黑名单”。黑白名单是所有反欺诈方法中最简单的，但也是更新最慢、成本最高的。

能将异常用户一网打尽

有监督学习需要大量有标签数据来训练模型，以此来预测还未被标注的数据。以垃圾邮件为例，假如把 5000 封已由人工确认过的垃圾邮件输入到模型，模型通过对标题的识别、邮件内容句子的分割、关键词的识别等各种分析方法，找到其中的内在关系。如标题中有“福利”二字的，有 90%的可能性是垃圾邮件；一次性发送超过 200 封的，有 60%的可能性是垃圾邮件；回复率低于 10%的，有 70%的可能性是垃圾邮件……于是，当模型处理一封新邮件时，通过检测以上各子项，并对每一子项乘以百分比后相加，就能得出垃圾邮件的可能性。但有监督学习的弊端是，每个模型都需要大量训练数据以及较长的训练时间。

“可能你的模型还没有训练好，欺诈分子已经完成欺诈活动并寻找下个目标了。”吴中说。

无监督机器学习主要方式有聚类和图形分析。金端峰说，无监督无需任何训练数据和标签，通过聚类等机器学习算法模型发现用户的共性行为，以及用户和用户的关系来检测欺诈。“通过无监督机器学习分析用户的共性行为，可以发现伪装过的异常用户，将其一网打尽。”

何为聚类方式？例如一群用户注册事件，可通过聚类发现几个小群符合某些共性：注册时间集中，都使用了某种操作系统，某一个浏览器版本等。该用户群中的任何一个单独拿出来分析，看上去都极为正常，如果符合某种超乎寻常的一致性就十分可疑了。比如一群人在凌晨 2—3 点采用同一款浏览器注册了同一产品，其 IP 的前 20 位相同，GPS 定位小于 1 公里，注册后都修改了昵称和性别等。

现在的金融欺诈都是团伙作战，面对“化整为零，批量复制”的欺诈手法，金端峰说，无监督算法应用于反欺诈检测还有一个优势，那就是能提前预警。“现在的欺诈分子都有潜伏期，以免太容易被发现。由于他们在潜伏期的行为依然符合某种规律，具有某些一致性，同样还是会被无监督算法捕捉到。在攻击发生前就检测出欺诈分子，这一点传统方法是难以做到的，防患于未然这也是无监督机器学习之所以在反欺诈检测中大放光彩的重要原因之一。”

防患于未然及时预警

在科技金融活动中，无监督机器学习能有效防止欺诈行为的发生并及时对用户发出预警，阻止开户欺诈、欺诈交易、账号盗取，发现洗钱攻击等，保障正常的金融活动。金端峰举例说，猛犸反欺诈公司基于非监督式的异常检测，将数据分解为正常趋势、随机扰动和异常情况三部分，并在此基础上做到设备、网络和用户三个层面上的“千人千面”；并根据用户间的相互关联构造网络图，欺诈者往往团体作案，行为表现在网络图中呈现高度一致性和聚集性，与正常用户明显不同，因此利用聚类和图形分析辨别欺诈行为。“蚂蚁金服、京东金融等一些高科技互联网公司也通过无监督机器学习等技术手段，在金融科技方面取得了良好成绩。”

除了有效防止欺诈行为的发生，无监督机器学习在科技金融领域还能有多种作用。比如通过用户画像和大数据模型精准找到用户，实现精准营销；根据个人投资者提供的风险承受水平、收益目标以及风格偏好等要求，运用一系列智能算法及投资组合优化等理论模型，为用户提供最终的投资参考，并依据市场动态对资产配置调整提供建议；投资研究需要收集大量资料，进行数据分析，报告撰写等，通过机器自主抓取相关信

息，可以辅助决策，甚至自动生成投研报告；利用大数据人工智能技术，可使用海量的多维度数据，塑造出高度精细化的风险控制模型；通过学习、积累金融法规，并结合金融机构的实际情况提供合规建议；机器还可以从海量的交易数据中学习知识和规则，发现异常行为，对洗钱行为进行警示等。

应用广泛可进行投资预测

无监督机器学习技术的应用正在不断深入和扩展。爱信诺是上市企业航天信息股份有限公司的全资子公司，在大数据采集、分析和应用方面具有突出能力，建成了以税务和企业经营数据为核心的企业信用数据库。

金端峰说，其实，许多大公司都有大型数据库，储存用户数据信息，通过无监督机器学习分析用户的整体数据，就能发现用户金融消费习惯的变化、投资偏好等，自动发现市场分类并针对不同群体用户推出不同的金融产品。“这样，有针对性的开发新市场，减少了盲目投入。”

此外，根据客户国籍、职业、薪酬、经验、行业、信用记录等信息，利用无监督机器学习技术来确定客户的信用风险评分，甚至是在向客户提供任何服务之前就进行此类评定，加快放贷过程，还能避免耗时而必要的“尽调”过程。

“随着机器学习的使用，股票预测变得相当简单。”金端峰说，机器学习算法会利用上市企业的资产负债表、损益表等历史数据，进行分析，并找出关系到公司未来发展的有意义的迹象，进行投资预测。