专利提案表（2018版）

**一、基本信息（加粗信息必填）**

|  |  |
| --- | --- |
| 提案名称：一种基于图特征的异常用户交易欺诈检测方法 | |
| 发明人: 黄勇 | |
| 承办人：黄勇 | 所属单位：电子支付研究院 |
| 座机: 20631812 | 手机: 13161788676 |
| 第一发明人: 黄勇 | 第一发明人身份证: 420321199206023812 |
| 涉及的项目（如有）： | 项目上线时间: |
| 共同专利申请人（如有）: | |

**★务必如实填写发明人信息,《专利法》规定，发明人（设计人）：指发明创造的实质性特点作出创造性贡献的人**

**二、构想的来源**

（什么原因引发了本专利提案的构思。如：（1）在完成项目的过程中，发现技术难点，为了克服技术难点。（2）在使用产品过程中发现产品的缺陷，为了克服缺陷。（3）对竞争对手或行业热点技术检索分析后进行的技术补充或者完善。（4）与技术标准相关的技术方案。）

近年来，随着移动支付的普及，越来越多的交易通过手机、智能手表等智能终端进行完成，大规模，高频词的交易行为为交易风控带来了巨大的挑战。同时，一些交易活动中也存在各种各样的异常用户，他们通过恶意的多次重复购买，占据资源，获取其他用户的福利，这些恶意用户也被叫做“黄牛”用户。异常交易和黄牛用户的存在，让金融交易中存在很多风险，如何及时发现交易中的风险，合理管理这些异常用户，也变得越来越重要。

如何识别交易中的欺诈用户和异常交易，发现在营销活动中的黄牛用户，无论对于公司还是对于正常的交易用户，都有着非常重要的作用。当前在交易欺诈检测中存在很多问题，**一是异常用户有较为相同的交易行为特征，**如图1是一些脱敏后的异常商户交易次数、交易金额特征，很容易发现这些商户的交易金额和交易次数很有规律，；**二是异常用户的交易行为和异常商户通常紧密关联**，一些交易终端如POS机、手机POS、二维码等和多个异常用户之间存在关联交易，这表明说明欺诈交易存在社群性，发现异常用户也可以通过关注POS机等收单设备的异常交易来进行挖掘，而当前研究并没有关注到商户侧的交易欺诈检测。

**表格 1 某营销活动中商户和终端异常交易统计图**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 终端 | 商户 | 最多出现的交易金额（元） | 最多出现次数 | 交易总量 |
| P1 | M1 | 5.00 | 4762 | 4771 |
| P2 | M1 | 40.00 | 2069 | 2083 |
| P3 | M2 | 80.00 | 2496 | 2703 |
| P4 | M3 | 79.00 | 2187 | 2389 |
| P5 | M4 | 80.00 | 2118 | 2155 |

本研究通过构建交易关联关系中的图特征模型，挖掘异常交易用户和商户。通过对商户、POS、用户构建图模型，并结合传统的异常交易欺诈检测模型，从而有效发现异常交易。

**三、如何发现他人侵权**

（请说明我司可以通过什么方式发现其他公司使用了本专利提案的技术方案。如：公开渠道、技术分析、码流分析、反向工程）

目前交易欺诈检测领域中可投入业界使用的异常用户检测系统，很多都是基于图特征构建的欺诈检测系统，不同系统使用的图特征模型具有很大的差异性。因此，若通过公开渠道发现相关的产品或方案类似的交易欺诈检测特征宣传，则有可能侵权。

**四、现有技术方案**

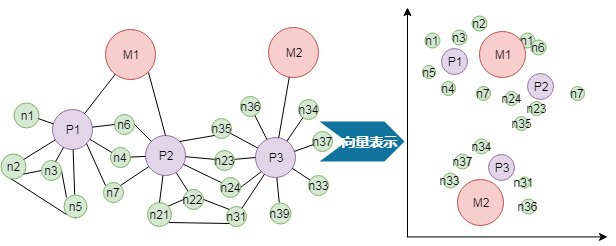
1、现有技术实现的方法/现有技术架构

（对所有相关现有技术方案作详细的描述，主要描述最接近的现有技术的现状，可以引用行业标准、行业规范、论文、技术标准、专利文献和实际产品。）

1. 交易欺诈

交易欺诈通常指一些用户为了获利，短时间内大量使用其他人的卡进行交易。目前交易欺诈的主要识别方法包括：（1）风控规则引擎，基于历史的业务经验和历史风控事件进行分析，根据业务不同、对风险的偏好和业务的处理不同，构建业务相关的风控规则，并发现交易欺诈；（2）异常检测方法，基于聚类或半监督的算法，对正常的交易和异常的交易进行聚类分析，从而发现交易中的异常点；（3）机器学习分类算法，通过交易特征因子、用户特征因子构建特征工程，利用神经网络算法、支持向量机、贝叶斯网络等训练分类模型对不同类型的交易进行分类识别，并预测未来相似模式的交易是否为异常用户欺诈交易。

1. 图特征表示

图特征模型对各种图（如社交网络、知识图谱、交易网络等）进行表示和建模，图特征模型通过将图中节点和边表示为特征向量来完成节点分类、边预测、节点预测等不同任务。如图1所示，将一个网络图转化为一个二维的向量。

**图 1 网络节点转化为二维向量表示示意图**

图特征模型可以分为四类不同的算法：（1）**基于矩阵分解的图特征算法**，这类算法将图中的节点和边转化为矩阵，对矩阵进行降维获取图中的节点和边的向量表示，这类算法能很好的获取图的全局结构关系；（2）**基于随机游走的图特征表示算法**，这类算法将图中一阶邻居节点、二阶邻居节点、高阶邻居节点作为目标节点的特征，借鉴自然语言预测模型，获取图中节点的向量表示，这类算法可以更好获取图中的局部特征，但算法复杂度较高；（3）**基于深度学习的图特征表示算法**，学习获取图中的节点和边的非线性结构特征，基于自动编码（autoencoder）等方法获取节点和边的向量表示；(4)**混合方法**，将矩阵分解、随机游走和深度学习不同类型的图特征表示方法结合，获取节点和边的特征表示。

2、现有技术存在的缺点或问题

（这里的缺点是本专利提案要解决的问题）

当前的交易欺诈模型主要缺点包括：（1）交易欺诈关注于用户发现，并没有综合考虑用户和商户，将用户和商户的交易网络进行建模；（2）当前的交易欺诈主要基于用户的行为特征进行分析，很少结合用户和商户的交易金额、行为特征、交易网络进行综合建模，从而预测异常的交易。

**五、本提案所述技术方案的技术效果**

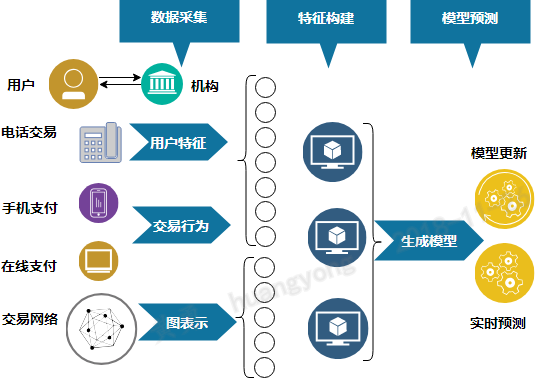
（按重要性顺序，列举并解释本专利提案的优点,技术效果需对应上述技术问题）

1. 与传统的交易欺诈检测相比，**本发明综合考虑商户和用户，将用户和商户组成的交易网络结合进行建模**。对整个交易关系网络进行表示学习建模，将用户节点和商户节点向量化建模，从而能对整个网络中的用户和商户节点进行分类，发现异常的用户和商户。
2. 本发明结合**用户的交易行为特征、用户的图表示向量特征、用户交易金额等多种类型特征进行综合建模**，能有效预测异常用户的交易欺诈，发现日常交易中的异常用户，营销活动中的欺诈交易。

**六、发明核心内容**

（简明扼要地描述整个技术方案/发明点，包含解决上述问题的必不可少的技术特征）

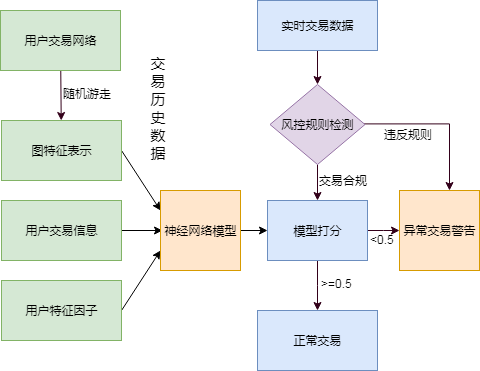
如图2是异常用户交易欺诈检测模型示意图,本发明提供一种基于图特征的异常用户交易欺诈检测方法。基于用户和商户的历史交易行为，构建用户（user）、交易设备（pos）、商户（merchant）交易网络，对交易网络中的用户和商户节点进行随机游走，构建节点的向量表示。对于用户的交易，获取用户交易的行为数据，并结合用户特征因子构建特征工程，建立神经网络模型。利用模型和风控规则对实时交易进行打分和规则判断，对异常的交易进行交易警告。



**图 2异常用户交易欺诈检测**

* 1. 对于用户的交易信息，包括用户使用的交易设备，交易位置，用户的购买商品及相关特征构建向量空间模型。
  2. 对于用户的特征因子，包括用户的职业、用户年龄、用户行为、用户兴趣爱好等信息构建成一个用户向量空间模型。
  3. 对于用户的交易网络，构建由用户、商户、支付设备、交易商品构建的交易网络，对网络中的节点进行随机游走，利用节点的一阶和二阶节点构建图特征表示，学习用户、商户和商品等不同节点的向量表示。
  4. 对于用户交易信息、用户特征因子和用户交易网络特征组合构建神经网络模型，对用户实时交易数据进行分类，判断是否有异常用户交易欺诈。

**七、最佳技术方案（实施例）**



**图 3异常用户交易欺诈检测流程图**

针对现有技术存在的不足和缺陷，本实例公开一种基于图特征的异常用户交易欺诈检测方法。如图三展示了异常用户交易欺诈检测流程图，图中基于历史交易数据构建用户交易网络，并基于随机游走获取用户和商户之间的图特征表示，结合用户交易信息和用户特征因子，训练神经网络模型。对于实时交易的数据，先通过风控规则检测，并对检测合规的交易再进行模型打分评价，评价符合规则则正常交易。若未通过风控规则检测或模型打分低于0.5，则进行异常交易警告，并按照业务方式进行相关处理。

**步骤一、用户交易网络构建**

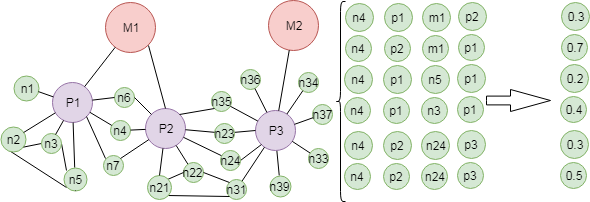
* 1. 用户交易数据获取。对于用户在线上交易、线下交易、二维码交易和闪付交易中的交易行为，获取相关的交易商户、交易设备号、交易商品等信息。
  2. 用户交易网络构建。对于交易记录中的用户、商户、交易设备号如POS机等进行信息抽取，构建交易网络。如通过构建三元组（u0001，交易，p0028）、（m0022,有设备,p0028）等三元组，则可构建商户、交易设备、用户的三层网络图。

**步骤二、图特征向量表示**

* 1. 用户交易网络映射。对于、用户ID、交易设备号、商户等构建字典进行映射转化为整数，并获取这些实体和关系对应的字符串和整数词典。
  2. 交易网络随机游走。如图3所示,对转化后的三元组构建图，对图中的每一个节点进行多次随机游走，获取随机游走后的序列，将这些序列作为该节点的特征表示。
  3. 交易图特征表示。对于随机游走获得的节点特征表示，构建深度学习模型，基于DeepWalk算法，利用网络节点的相邻节点预测，从而学习的每个节点的向量表示。

**步骤三、用户和行为数据收集**

* 1. 获取用户相关的特征数据，包括用户的卡相关信息、用户的个人相关特征因子，如个人的职业、消费标签等信息。
  2. 获取交易行为相关的数据，包括交易的地点、交易的IP地址、交易金额等交易数据进行处理。
  3. 将上述的用户数据和交易行为数据进行处理，包括对职业信息进行哈希编码，字符特征因子转化为数字向量,分别记为和。



**图 4 随机游走计算节点向量表示**

**步骤四、神经网络模型构建**

* 1. 对上述步骤一、步骤二和步骤三获取的特征因子进行组合，获取联合向量特征因子记为V=[]，将联合向量特征因子作为输入向量，构建神经网络分类模型，对交易数据进行分类，发现其中的异常交易。
  2. 构建神经网络模型进行训练。对于有标记的历史交易数据,构建多层神经网络,进行分类器的模型训练。其中=[],表示第i笔历史交易数据，N表示训练历史数据的总数量，=，表示数据是否被标记为欺诈交易。
  3. 神经网络模型的输向量是[], 隐藏层，输出层是，其中采用ReLU激活函数，为第l层的学习权重，为偏置，为激活函数学习得到第l+1层的值，此外对不同的数据，为了能获得更好的预测结果，激活函数也可以采用softmax、sigmoid等其他激活函数。

**步骤五、实时交易风控评估**

* 1. 对于线上的实时交易数据，进行实时判断。首先对数据判断是否符合公司内部的业务规则，若符合规则则使用神经网络模型继续判断，否则对该交易发出警告，进行其他操作。
  2. 对符合业务规则的数据进行模型预测，判断该交易数据是否在结合用户特征、交易行为和交易图特征的神经网络模型中是否可能是正常交易。若超过设定阈值判断为正常交易，否则认定为异常交易，发出警告并进行其他操作。

**八、可替换技术方案、补充实施例**

（描述可替换上述技术方案的其他技术方案。例如：（1）修改该发明，也能得到实质性同样或接近的结果；（2）方法流程顺序有变化，或者有新增步骤加入；（2）装置方案中有组件替换，或者结构变形等。）

1. 上述用户交易网络构建基于“商户-设备号-用户”三层的网络图进行构建，如果可以获取用户或商户其他的关系，如增加用户间的转账关系，用户间的朋友关系等，基于更多不同类型的关系构建网络图，则能提高图特征表示建模的真实性，提高网络中节点向量表示的真实性。其中的用户和商户之间的交易关系也可以替换为其他的交易关系；
2. 上述的基于DeepWalk的随机游走算法可以获取节点的向量表示，也可以替换为其他的图特征向量表示算法，如node2vec、LINE、TransE等不同的表示图学习方法。
3. 上述的神经网络模型基于多层神经网络模型构建，也可以采用CNN、RNN、Random Forest、GBDT等不同的监督学习算法进行替换。
4. 上述的多层神经网络采用随机梯度优化算法学习模型的参数，可替换为其他的不同种类梯度学习算法；
5. 上述的损失函数采用对数损失函数，可替换为均方误差、边际损失等其他类型的损失函数；
6. 上述的激活函数采用ReLU激活函数，根据预测模型训练效果的不同，可以采用softmax、sigmoid等激活函数；
7. 上述的模型评价指标采用AUC进行评价，可替换为其他基于分类或排序算法使用的损失函数如MAP、f1等指标；

**九、术语说明表**

（请将本提案涉及的技术术语按照下表形式填写，尤其是发明点相关的技术术语和本发明自定义的术语）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 英文缩写 | 英文全称 | 中文全称 | 中文解释 | 其他说明 |
| DeepWalk | DeepWalk |  | 基于随机游走算法，以预测图中的相邻节点为目标，学习节点向量表示算法。 |  |
| LINE | Large-scale Information Network Embedding |  | 基于图中的一阶邻居节点和二阶邻居节点计算图中界定啊向量表示 |  |
| Random Walk | Random Walk | 随机游走算法 | 发现图中节点和节点之间关系的算法 | 给定图和出发点，随机选择邻居结点，移动到邻居结点，并把新结点作为出发点重复游走，获取关系路径 |