

权利要求书

- 1、一种分类模型生成方法，其特征在于，包括：

获取训练样本数据集，其中，所述训练样本数据集中包括至少两个样本数据以及每个所述样本数据对应的样本标签；
- 5 构建所述至少两个样本数据的第一关系网络，通过所述第一关系网络的节点表征所述至少两个样本数据；

基于预设的网络嵌入学习模型获得所述第一关系网络中每个节点的向量化表示；

通过所述训练样本数据集对分类模型进行训练，得到所述分类模型，所述分类模型使得所述样本标签与所述第一关系网络中每个节点的向量化表示相关联。
- 10 2、根据权利要求1所述的方法，其特征在于，获取训练样本数据集包括：

按照预设时间间隔获取训练样本数据集。
- 3、根据权利要求1所述的方法，其特征在于，基于预设的网络嵌入学习模型获得所述第一关系网络中每个节点的向量化表示包括：

采用随机游走算法对所述第一关系网络中每个节点进行序列采样，并生成第一节点
- 15 序列；

基于预设的网络嵌入学习模型将所述第一节点序列中的每个节点进行向量化表示。
- 4、根据权利要求3所述的方法，其特征在于，基于预设的网络嵌入学习模型将所述第一节点序列中的每个节点进行向量化表示包括：

基于 Node2vec 网络嵌入学习模型将所述第一节点序列中的每个节点进行向量化表
- 20 示。
- 5、根据权利要求3所述的方法，其特征在于，基于预设的网络嵌入学习模型将所述第一节点序列中的每个节点进行向量化表示包括：

基于 DeepWalk 网络嵌入学习模型将所述第一节点序列中的每个节点进行向量化表示。
- 25 6、根据权利要求4所述的方法，其特征在于，基于 Node2vec 网络嵌入学习模型将所述第一节点序列中的每个节点进行向量化表示包括：

所述 Node2vec 网络嵌入学习模型基于 Word2vec 的 SkipGram 框架将所述节点序列中的每个节点进行向量化表示。
- 7、根据权利要求1所述的方法，其特征在于，所述分类模型包括 XGBoost 模型、
- 30 随机森林模型、支持向量机模型或逻辑回归模型。
- 8、一种数据识别方法，其特征在于，包括：

获取待识别数据集，其中，所述待识别数据集包括至少两个待识别数据；

构建所述至少两个待识别数据的第二关系网络，通过所述第二关系网络的节点表征所述至少两个待识别数据；

基于预设的网络嵌入学习模型获得所述第二关系网络中每个节点的向量化表示；

5 根据预先训练的分类模型确定所述第二关系网络中每个节点的向量化表示对应的数据标签。

9、根据权利要求 8 所述的方法，其特征在于，基于预设的网络嵌入学习模型获得所述第二关系网络中每个节点的向量化表示包括：

10 采用随机游走算法对所述第二关系网络中每个节点进行序列采样，并生成第二节点序列；

根据预设的网络嵌入学习模型将所述第二节点序列中的每个节点进行向量化表示。

10、根据权利要求 9 所述的方法，其特征在于，根据预设的网络嵌入学习模型将所述第二节点序列中的每个节点进行向量化表示包括：

15 基于 Node2vec 网络嵌入学习模型将所述第二节点序列中的每个节点进行向量化表示。

11、根据权利要求 9 所述的方法，其特征在于，根据预设的网络嵌入学习模型将所述第二节点序列中的每个节点进行向量化表示包括：

基于 DeepWalk 网络嵌入学习模型将所述第二节点序列中的每个节点进行向量化表示。

20 12、根据权利要求 10 所述的方法，其特征在于，基于 Node2vec 网络嵌入学习模型将所述第二节点序列中的每个节点进行向量化表示包括：

所述 Node2vec 网络嵌入学习模型基于 Word2vec 的 SkipGram 框架将所述第二节点序列中的每个节点进行向量化表示。

25 13、根据权利要求 8 所述的方法，其特征在于，所述数据标签包括白数据标签和黑数据标签。

14、根据权利要求 13 所述的方法，其特征在于，还包括：

若所述数据标签为黑数据标签，则对所述数据标签对应的所述第二关系网络中节点的向量化表示进行后续处理。

30 15、根据权利要求 8 所述的方法，其特征在于，所述待识别数据集包括通过所述分类模型判断是否存在洗钱行为的待识别账户的集合。

16、一种分类模型生成装置，其特征在于，包括：

第一获取模块,被配置为获取训练样本数据集,其中,所述训练样本数据集中包括至少两个样本数据以及每个所述样本数据对应的样本标签;

第一构建模块,被配置为构建所述至少两个样本数据的第一关系网络,通过所述第一关系网络的节点表征所述至少两个样本数据;

- 5 第一学习模块,被配置为基于预设的网络嵌入学习模型获得所述第一关系网络中每个节点的向量化表示;

训练模块,被配置为通过所述训练样本数据集对分类模型进行训练,得到所述分类模型,所述分类模型使得所述样本标签与所述第一关系网络中每个节点的向量化表示相关联。

- 10 17、根据权利要求 16 所述的装置,其特征在于,所述第一学习模块包括:

第一生成子模块,被配置为采用随机游走算法对所述第一关系网络中每个节点进行序列采样,并生成第一节点序列;

第二学习子模块,被配置为基于预设的网络嵌入学习模型将所述第一节点序列中的每个节点进行向量化表示。

- 15 18、一种数据识别装置,其特征在于,包括:

第二获取模块,被配置为获取待识别数据集,其中,所述待识别数据集包括至少两个待识别数据;

第二构建模块,被配置为构建所述至少两个待识别数据的第二关系网络,通过所述第二关系网络的节点表征所述至少两个待识别数据;

- 20 第三学习模块,被配置为基于预设的网络嵌入学习模型获得所述第二关系网络中每个节点的向量化表示;

识别模块,被配置为根据预先训练的分类模型确定所述第二关系网络中每个节点的向量化表示对应的数据标签。

- 19、根据权利要求 18 所述的装置,其特征在于,所述第三学习模块包括:

- 25 第二生成子模块,被配置为采用随机游走算法对所述第二关系网络中每个节点进行序列采样,并生成第二节点序列;

第四学习子模块,被配置为根据预设的网络嵌入学习模型将所述第二节点序列中的每个节点进行向量化表示。

- 30 20、一种计算设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机指令,其特征在于,所述处理器执行所述指令时实现权利要求 1-7 或 8-15 任意一项所述方法的步骤。

21、一种计算机可读存储介质，其存储有计算机指令，其特征在于，该程序被处理器执行时实现权利要求 1-7 或 8-15 任意一项所述方法的步骤。