

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 8 月 6 日 (06.08.2020)



(10) 国际公布号
WO 2020/155831 A1

(51) 国际专利分类号:
G06N 20/00 (2019.01)

(21) 国际申请号: PCT/CN2019/121902

(22) 国际申请日: 2019 年 11 月 29 日 (29.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910095815.5 2019 年 1 月 31 日 (31.01.2019) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。

(72) 发明人: 程羽 (CHENG, Yu); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。刘腾飞 (LIU, Tengfei); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。王维强 (WANG, Weiqiang); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。杨洋 (YANG, Yang); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。徐轶 (XU, Yi); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市

(54) Title: DATA TAG GENERATION METHOD AND APPARATUS, MODEL TRAINING METHOD AND APPARATUS, AND EVENT IDENTIFICATION METHOD AND APPARATUS

(54) 发明名称: 数据标签生成、模型训练、事件识别方法和装置

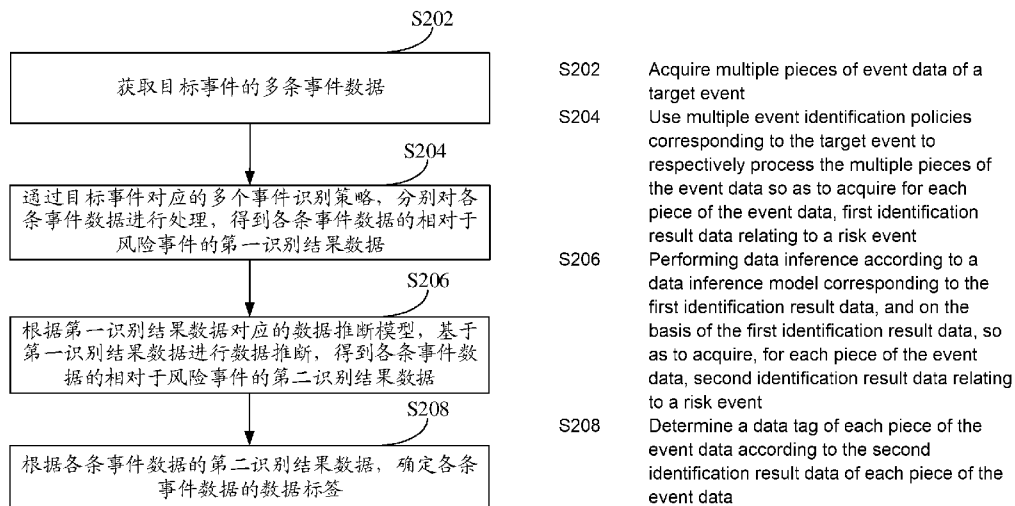


图 2

(57) Abstract: A data tag generation method and apparatus, a model training method and apparatus, and an event identification method and apparatus. The data tag generation method comprises: acquiring multiple pieces of event data of a target event (S202); using multiple event identification policies corresponding to the target event to respectively process the multiple pieces of the event data so as to acquire, for each piece of the event data, first identification result data relating to a risk event (S204); performing data inference according to a data inference model corresponding to the first identification result data, and on the basis of the first identification result data, so as to acquire, for each piece of the event data, second identification result data relating to a risk event (S206); and determining a data tag of each piece of the event data according to the second identification result data of each piece of the event data (S208).



海淀区上地三街9号嘉华大厦B座409,
Beijing 100085 (CN)。

- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。

(57) 摘要: 一种数据标签生成、模型训练、事件识别方法和装置, 其中数据标签生成方法包括: 获取目标事件的多条事件数据(S202); 通过所述目标事件对应的多个事件识别策略, 分别对各条所述事件数据进行处理, 得到各条所述事件数据的相对于风险事件的第一识别结果数据(S204); 根据所述第一识别结果数据对应的数据推断模型, 基于所述第一识别结果数据进行数据推断, 得到各条所述事件数据的相对于风险事件的第二识别结果数据(S206); 根据各条所述事件数据的所述第二识别结果数据, 确定各条所述事件数据的数据标签(S208)。

数据标签生成、模型训练、事件识别方法和装置

技术领域

[01]本申请涉及计算机设备领域，尤其涉及一种数据标签生成、模型训练、事件识别方法和装置。

背景技术

[02]随着人工智能技术和深度学习技术的发展，可以在计算机内建立各种模型，通过模型对各种事件进行处理。比如，通过用户投诉事件识别模型对用户投诉事件进行处理，以判断用户投诉事件是否成立。用户投诉事件可以是投诉其他用户赌博、洗钱等事件。

[03]目前训练模型的主要过程为：获取样本数据，对样本数据打标签，利用打标签后的样本数据训练模型。目前主要通过人工手动的方式对数据打标签，存在标签的准确性难以保证的问题，导致模型的数据处理结果准确率低。

发明内容

[04]本申请实施例的目的是提供一种数据标签生成、模型训练、事件识别方法和装置，以解决人工手动打标签存在的标签准确性难以保证的问题，提高模型的数据处理结果的准确率。

[05]为解决上述技术问题，本申请实施例是这样实现的：

[06]本申请实施例提供了一种数据标签生成方法，包括：

[07]获取目标事件的多条事件数据；

[08]通过所述目标事件对应的多个事件识别策略，分别对各条所述事件数据进行处理，得到各条所述事件数据的相对于风险事件的第一识别结果数据；

[09]根据所述第一识别结果数据对应的数据推断模型，基于所述第一识别结果数据进行数据推断，得到各条所述事件数据的相对于风险事件的第二识别结果数据；

[10]根据各条所述事件数据的所述第二识别结果数据，确定各条所述事件数据的数据标签。

[11]本申请实施例提供了一种模型训练方法，包括：

[12]根据上述的数据标签生成方法，确定目标事件的事件数据的数据标签；

[13]将所述目标事件的事件数据作为所述目标事件的样本数据；

[14]根据所述目标事件的样本数据和所述样本数据的数据标签，训练所述目标事件对应的风险事件识别模型，其中，所述风险事件识别模型用于识别所述目标事件是否为风险事件，或者，用于识别所述目标事件为风险事件的概率。

[15]本申请实施例提供了一种事件识别方法，包括：

[16]获取待识别目标事件的事件数据；

[17]利用上述模型训练方法训练得到的风险事件识别模型，对所述待识别目标事件的事件数据进行处理；

[18]根据处理结果，确定所述待识别目标事件是否为风险事件，或者，确定所述待识别目标事件为风险事件的概率。

[19]本申请实施例提供了一种数据标签生成装置，包括：

[20]第一获取模块，用于获取目标事件的多条事件数据；

[21]第一结果确定模块，用于通过所述目标事件对应的多个事件识别策略，分别对各条所述事件数据进行处理，得到各条所述事件数据的相对于风险事件的第一识别结果数据；

[22]第二结果确定模块，用于根据所述第一识别结果数据对应的数据推断模型，基于所述第一识别结果数据进行数据推断，得到各条所述事件数据的相对于风险事件的第二识别结果数据；

[23]标签生成模块，用于根据各条所述事件数据的所述第二识别结果数据，确定各条所述事件数据的数据标签。

[24]本申请实施例提供了一种模型训练装置，包括：

[25]标签确定模块，用于根据上述的数据标签生成装置，确定目标事件的事件数据的数据标签；

[26]样本确定模块，用于将所述目标事件的事件数据作为所述目标事件的样本数据；

[27]模型训练模块，用于根据所述目标事件的样本数据和所述样本数据的数据标签，训练所述目标事件对应的风险事件识别模型，其中，所述风险事件识别模型用于识别所述目标事件是否为风险事件，或者，用于识别所述目标事件为风险事件的概率。

[28]本申请实施例提供了一种事件识别装置，包括：

[29]第二获取模块，用于获取待识别目标事件的事件数据；

[30]数据处理模块，用于利用上述的模型训练装置训练得到的风险事件识别模型，对所述待识别目标事件的事件数据进行处理；

[31]事件识别模块，用于根据处理结果，确定所述待识别目标事件是否为风险事件，或者，确定所述待识别目标事件为风险事件的概率。

[32]本申请实施例提供了一种电子设备，包括：处理器；以及被安排成存储计算机可执行指令的存储器，所述计算机可执行指令在被执行时使所述处理器实现上述的数据标签生成方法的步骤，或者，实现上述的模型训练方法的步骤，或者，实现上述的事件识别方法的步骤。

[33]本申请实施例提供了一种存储介质，用于存储计算机可执行指令，所述计算机可执行指令在被执行时实现上述的数据标签生成方法的步骤，或者，实现上述的模型训练方法的步骤，或者，实现上述的事件识别方法的步骤。

[34]本实施例中，首先获取目标事件的多条事件数据，并通过目标事件对应的多个事件识别策略，分别对各条事件数据进行处理，得到各条事件数据的相对于风险事件的第一识别结果数据，然后根据第一识别结果数据对应的数据推断模型，基于第一识别结果数据进行数据推断，得到各条事件数据的相对于风险事件的第二识别结果数据，最后根据各条事件数据的第二识别结果数据，确定各条事件数据的数据标签。本实施例中，通过数据推断的方式，能够准确得到各条事件数据的识别结果数据，从而准确确定事件数据的数据标签，解决人工手动打标签存在的标签准确性难以保证的问题，提高数据打标签的准确率，提高模型的数据处理结果的准确率。

附图说明

[35]为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请中记载的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

[36]图1为本申请一实施例提供的本实施例中的各个方法的应用场景示意图；

[37]图2为本申请一实施例提供的数据标签生成方法的流程示意图；

[38]图 3 为本申请另一实施例提供的数据标签生成方法的流程示意图；

[39]图 4 为本申请另一实施例提供的数据标签生成方法的流程示意图；

[40]图 5 为本申请一实施例提供的模型训练方法的流程示意图；

[41]图 6 为本申请一实施例提供的事件识别方法的流程示意图；

[42]图 7 为本申请一实施例提供的一种数据标签生成装置的模块组成示意图；

[43]图 8 为本申请一实施例提供的一种模型训练装置的模块组成示意图；

[44]图 9 为本申请一实施例提供的一种事件识别装置的模块组成示意图；

[45]图 10 为本申请一实施例提供的电子设备的结构示意图。

具体实施方式

[46]为了使本技术领域的人员更好地理解本申请中的技术方案，下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都应当属于本申请保护的范围。

[47]本申请实施例提供了一种数据标签生成、模型训练、事件识别方法和装置，以解决人工手动打标签存在的标签准确性难以保证的问题，提高模型的数据处理结果的准确率。

[48]图 1 为本申请一实施例提供的本实施例中的各个方法的应用场景示意图，如图 1 所示，图 1 提供了一种计算机系统，该系统包括数据标签生成子系统 100、模型训练子系统 200 和事件识别子系统 300，其中，数据标签生成子系统 100 可以执行本实施例中的数据标签生成方法，以为各条事件数据打标签。模型训练子系统 200 可以执行本实施例中的模型训练方法，基于打标签完成的各条事件数据，训练风险事件识别模型，风险事件识别模型用于识别目标事件是否为风险事件，或者，用于识别目标事件为风险事件的概率。事件识别子系统 300 可以执行本实施例中的事件识别方法，运行模型训练子系统 200 训练得到的风险事件识别模型，从而识别目标事件是否为风险事件，或者，识别目标事件为风险事件的概率。

[49]本实施例中，图 1 中的数据标签生成子系统 100、模型训练子系统 200 和事件识别子系统 300 可以分别包括一个或多个计算机设备，比如，通过若干个计算机设备组成数

据标签生成子系统 100, 通过若干个计算机设备组成模型训练子系统 200, 通过若干个计算机设备组成事件识别子系统 300。

[50]在其他实施例中, 也可以通过两个子系统实现本实施例中的各个方法, 比如, 通过一个子系统执行数据标签生成方法和模型训练方法, 通过另一个子系统执行事件识别方法, 或者, 通过一个子系统执行数据标签生成方法, 通过另一个子系统执行模型训练方法和事件识别方法。其中, 每个子系统分别包括一个或多个计算机设备。

[51]在其他实施例中, 也可以通过一个计算机设备实现上述的数据标签生成方法、模型训练方法和事件识别方法。需要说明的是, 以上提及的各种实施例所达到的功能和效果均相同, 这里不再重复。

[52]图 2 为本申请一实施例提供的数据标签生成方法的流程示意图, 如图 2 所示, 该流程包括以下步骤:

[53]步骤 S202, 获取目标事件的多条事件数据;

[54]步骤 S204, 通过目标事件对应的多个事件识别策略, 分别对各条事件数据进行处理, 得到各条事件数据的相对于风险事件的第一识别结果数据;

[55]步骤 S206, 根据第一识别结果数据对应的数据推断模型, 基于第一识别结果数据进行数据推断, 得到各条事件数据的相对于风险事件的第二识别结果数据;

[56]步骤 S208, 根据各条事件数据的第二识别结果数据, 确定各条事件数据的数据标签。

[57]本实施例中, 首先获取目标事件的多条事件数据, 并通过目标事件对应的多个事件识别策略, 分别对各条事件数据进行处理, 得到各条事件数据的相对于风险事件的第一识别结果数据, 然后根据第一识别结果数据对应的数据推断模型, 基于第一识别结果数据进行数据推断, 得到各条事件数据的相对于风险事件的第二识别结果数据, 最后根据各条事件数据的第二识别结果数据, 确定各条事件数据的数据标签。本实施例中, 通过数据推断的方式, 能够准确得到各条事件数据的识别结果数据, 从而准确确定事件数据的数据标签, 解决人工手动打标签存在的标签准确性难以保证的问题, 提高数据打标签的准确率, 提高模型的数据处理结果的准确率。

[58]上述步骤 S202 中, 目标事件可以为用户交易事件, 目标事件的事件数据可以为用户交易事件对应的交易数据。比如, 用户通过电商平台进行多次商品购买活动, 则目标事件设定为用户的买卖交易事件, 目标事件的事件数据设定为用户的买卖交易事件对应的交易数据。本实施例中, 交易数据包括但不限于交易时间、交易金额、交易时的网络

环境参数、交易所用的终端设备的设备参数、交易金额扣款途径、商家信息、收款账户信息等。

[59]一个具体的实施例中，选取用户的信用卡交易事件为目标事件，并设置风险事件为套现事件，从而为用户的信用卡交易数据打标签，该标签可以用于表示对应的信用卡交易事件为套现事件的概率。另一个具体的实施例中，选取用户的购物交易事件为目标事件，并设置风险事件为赌博事件，从而为用户的购物交易数据打标签，该标签可以用于表示对应的购物交易事件为赌博事件的概率。

[60]本实施例中，预设有风险数据仓库，该仓库中存储有大量用户的交易数据，步骤 S202 中，可以从该风险数据仓库中获取多条交易数据作为事件数据。

[61]本实施例中，预先配置有多个事件识别策略，事件识别策略可以对目标事件的事件数据进行处理，以识别目标事件是否为风险事件，或者识别目标事件为风险事件的概率。以目标事件为信用卡交易事件，风险事件为套现事件为例，通过事件识别策略可以识别信用卡交易事件是否为套现事件。事件识别策略可以包括各种模型或规则等。

[62]上述步骤 S204 中，通过目标事件对应的多个事件识别策略，分别对各条事件数据进行处理，得到各条事件数据的相对于风险事件的第一识别结果数据。由于分别通过多个事件识别策略对各条事件数据进行处理，因此每条事件数据均具有多个第一识别结果数据。一个实施例中，每条事件数据通过一个事件识别策略识别得到一个第一识别结果数据，每条事件数据的第一识别结果数据的数量与事件识别策略的数量相同。

[63]在一个具体的实施例中，事件识别策略包括模型和规则。通过模型可以对事件数据进行处理，从而确定目标事件为风险事件的概率，通过规则（比如基于 If-Then 的规则）可以对事件数据进行处理，从而确定目标事件是否为风险事件。其中，模型和规则中涉及的各个阈值可以基于人工经验确定。本实施例中，若模型或规则的输出结果为分值形式，则可以对模型或规则输出的各个分值进行归一化处理，比如将分值都归一化到 0-10 分之间，从而方便后续处理。

[64]由于通过事件识别策略识别得到的第一识别结果数据可能不太准确，基于第一识别结果数据为事件数据打标签可能存在准确率低的问题，因此本实施例中第一识别结果数据又可以称为事件数据的“弱标签”。

[65]下表 1 为本申请一实施例提供的第一识别结果数据的示意表，如表 1 所示，以三个事件识别策略、四条事件数据为例，通过每个事件识别策略，能够为每条事件数据识别

得到一个第一识别结果数据，从而得到下表 1。

表 1

	事件识别策略 1	事件识别策略 2	事件识别策略 3
第一条事件数据	第一识别结果数据 1.1	第一识别结果数据 2.1	第一识别结果数据 3.1
第二条事件数据	第一识别结果数据 1.2	第一识别结果数据 2.2	第一识别结果数据 3.2
第三条事件数据	第一识别结果数据 1.3	第一识别结果数据 2.3	第一识别结果数据 3.3
第四条事件数据	第一识别结果数据 1.4	第一识别结果数据 2.4	第一识别结果数据 3.4

[66]在得到多个第一识别结果数据后，本实施例中，可以确定第一识别结果数据对应的数据推断模型，具体确定方式如下：

[67]（a1）在得到各条事件数据的相对于风险事件的第二识别结果数据之前，将每个事件识别策略的针对各条事件数据的多个第一识别结果数据确定为一组第一识别结果数据；

[68]（a2）对各组第一识别结果数据进行分析，以确定各组第一识别结果数据之间的数据相关性分析结果；

[69]（a3）分别对各组第一识别结果数据进行分布统计，以确定各组第一识别结果数据各自对应的分布数据；其中，第一识别结果数据可以为分值数据；

[70]（a4）根据数据相关性分析结果和/或分布数据，确定第一识别结果数据对应的数据推断模型，数据推断模型包括投票模型、概率图模型和矩阵分解模型中的至少一种。

[71]由表 1 可知，每个事件识别策略基于多条事件数据对应一组第一识别结果数据，比如表 1 中的第一识别结果数据 1.1、第一识别结果数据 1.2、第一识别结果数据 1.3 和第一识别结果数据 1.4 组成一组第一识别结果数据。因此上述动作（a1）中，在得到各条事件数据的相对于风险事件的第二识别结果数据之前，将每个事件识别策略的针对各条事件数据的多个第一识别结果数据确定为一组第一识别结果数据，表 1 中共有三组第一识别结果数据。

[72]上述动作（a2）中，对各组第一识别结果数据进行分析，以确定各组第一识别结果数据之间的数据相关性分析结果，比如，将表 1 中从左至右的每列第一识别结果数据确定为一组第一识别结果数据，对表 1 中的这三组第一识别结果数据进行分析，以确定任

意两组第一识别结果数据之间的数据相关性。数据相关性是指数据之间存在某种关系，如正相关，负相关等，具体分析方法可以采用通用的分析算法，这里不做限定。

[73]上述动作(a3)中，分别对各组第一识别结果数据进行分布统计，以确定各组第一识别结果数据各自对应的分布数据。一个具体的实施例，分别对各组第一识别结果数据进行分布统计，得到各组第一识别结果数据各自对应的数据分布函数，将该数据分布函数作为各组第一识别结果数据各自对应的分布数据。当然，该分布数据也可以为数据分布直方图或者数据分布曲线图等，这里不做限定。

[74]上述动作(a4)中，根据数据相关性分析结果和/或分布数据，确定第一识别结果数据对应的数据推断模型，数据推断模型包括投票模型、概率图模型和矩阵分解模型中的至少一种。

[75]一个实施例中，可以人工根据数据相关性分析结果和/或分布数据，在投票模型、概率图模型和矩阵分解模型中选择一种模型作为第一识别结果数据对应的数据推断模型。

[76]另一个实施例中，若数据相关性分析结果表示各组第一识别结果数据之间高度相关，则可以选择概率图模型或矩阵分解模型作为数据推断模型，若数据相关性分析结果表示各组第一识别结果数据之间不相关，则可以选择投票模型作为数据推断模型。

[77]另一个实施例中，若上述分布数据满足特定的分布形态要求，则可以选择概率图模型或矩阵分解模型作为数据推断模型，若上述分布数据不满足特定的分布形态要求，则可以选择投票模型作为数据推断模型。

[78]另一个实施例中，若数据相关性分析结果表示各组第一识别结果数据之间高度相关，且上述分布数据满足特定的分布形态要求，则可以选择概率图模型或矩阵分解模型作为数据推断模型，若数据相关性分析结果表示各组第一识别结果数据之间不相关，或者，上述分布数据不满足特定的分布形态要求，则可以选择投票模型作为数据推断模型。

[79]根据数据相关性分析结果和/或分布数据在投票模型、概率图模型和矩阵分解模型中选择一种模型作为数据推断模型的方式有多种，这里不再一一列举。

[80]在确定数据推断模型后，可以执行上述步骤 S206，根据第一识别结果数据对应的数据推断模型，基于第一识别结果数据进行数据推断，得到各条事件数据的相对于风险事件的第二识别结果数据，该过程具体为：通过所述数据推断模型，对各条事件数据对应的各个第一识别结果数据进行运算，得到各条事件数据的相对于风险事件的第二识别结果数据。

[81]具体地，以数据推断模型为投票模型为例，结合上述表 1，本步骤中可以通过投票模型对每条事件数据对应的各个第一识别结果数据进行投票运算，得到每条事件数据的相对于风险事件的第二识别结果数据。其中，第二识别结果数据可以是各个第一识别结果数据中投票选择确定的，也可以是重新生成的识别结果数据。

[82]以数据推断模型为概率图模型为例，结合上述表 1，本步骤中可以通过概率图模型以每条事件数据对应的各个第一识别结果数据为基础建立概率图，根据该概率图确定每条事件数据的相对于风险事件的第二识别结果数据。其中，第二识别结果数据可以是各个第一识别结果数据中选择确定的，也可以是重新生成的识别结果数据。

[83]通过矩阵分解模型对每条事件数据对应的各个第一识别结果数据进行运算，得到每条事件数据的相对于风险事件的第二识别结果数据的过程可以参考矩阵分解模型的具体算法过程，这里不再重复。

[84]由于本实施例中通过数据推断的方式对每条事件数据对应的各个第一识别结果数据进行运算，得到每条事件数据的相对于风险事件的第二识别结果数据，因此得到的第二识别结果数据具有准确性高的优点。

[85]上述步骤 S208 中，根据各条事件数据的第二识别结果数据，确定各条事件数据的数据标签，可以为：将各条事件数据的第二识别结果数据，分别确定为各条事件数据的数据标签，比如第一条事件数据的数据标签为 1，表示该条事件数据对应的目标事件为赌博事件，第二条事件数据的数据标签为 0，表示该条事件数据对应的目标事件不为赌博事件，其中，赌博事件即为前述的风险事件。

[86]通过以上过程，能够为各条事件数据确定准确的数据标签，从而解决人工手动打标签存在的标签准确性难以保证的问题，提高数据打标签的准确率，提高模型的数据处理结果的准确率。

[87]考虑到通过步骤 S204 得到各条事件数据的相对于风险事件的第一识别结果数据后，可能存在第一识别结果数据为缺失数据，或者第一识别结果数据过于不准确不具有参考依据的情况，本实施例中的方法在得到各条事件数据的相对于风险事件的第二识别结果数据之前，还可以：根据第一识别结果数据，在各个事件识别策略中确定问题策略，在第一识别结果数据中删除问题策略对应的第一识别结果数据。相应地，根据第一识别结果数据对应的数据推断模型，基于第一识别结果数据进行数据推断，具体为：根据删除后的第一识别结果数据对应的数据推断模型，基于删除后的第一识别结果数据进行数据

推断。

[88] 其中，根据第一识别结果数据，在各个事件识别策略中确定问题策略，可以为：

[89] (b1) 将每个事件识别策略的针对各条事件数据的多个第一识别结果数据确定为一组第一识别结果数据；

[90] (b2) 对第一识别结果数据分别执行以下处理中的一项或多项，根据执行结果，在各个事件识别策略中确定问题策略；

[91] (b21) 分别统计各组第一识别结果数据中表示识别结果为空缺的第一识别结果数据的数量比例；

[92] (b22) 分别对各组第一识别结果数据进行分布统计，以确定各组第一识别结果数据各自对应的分布数据；其中，第一识别结果数据可以为分值数据；

[93] (b23) 对各组第一识别结果数据进行分析，以确定各组第一识别结果数据之间的数据相关性分析结果。

[94] 由表 1 可知，每个事件识别策略基于多条事件数据对应一组第一识别结果数据，比如表 1 中的第一识别结果数据 1.1、第一识别结果数据 1.2、第一识别结果数据 1.3 和第一识别结果数据 1.4 组成一组第一识别结果数据。因此上述动作 (b1) 中，将每个事件识别策略的针对各条事件数据的多个第一识别结果数据确定为一组第一识别结果数据，表 1 中共有三组第一识别结果数据。

[95] 上述动作 (b21) 中，分别统计各组第一识别结果数据中表示识别结果为空缺的第一识别结果数据的数量比例。以表 1 为例，统计表 1 中每列数据（即每组第一识别结果数据）中表示识别结果为空缺的第一识别结果数据的数量比例。在通过事件识别策略对目标事件的事件数据进行处理时，事件识别策略的识别结果数据可能表示识别结果为空缺，即未得到识别结果的情况，基于此，本实施例中统计空缺结果的数量比例，若空缺结果的数量比例偏高，则说明对应的事件识别策略不适用于处理该目标事件的事件数据。

[96] 上述动作 (b22) 中，分别对各组第一识别结果数据进行分布统计，以确定各组第一识别结果数据各自对应的分布数据。一个具体的实施例中，分别对各组第一识别结果数据进行分布统计，得到各组第一识别结果数据各自对应的数据分布函数，将该数据分布函数作为各组第一识别结果数据各自对应的分布数据。当然，该分布数据也可以为数据分布直方图或者数据分布曲线图等，这里不做限定。

[97]上述动作(b23)中,对各组第一识别结果数据进行分析,以确定各组第一识别结果数据之间的数据相关性分析结果,比如,将表1中从左至右的每列第一识别结果数据确定为一组第一识别结果数据,对表1中的这三组第一识别结果数据进行分析,以确定任意两组第一识别结果数据之间的数据相关性,或者确定任意三组第一识别结果数据之间的数据相关性。数据相关性是指数据之间存在某种关系,如正相关,负相关等,具体分析方法可以采用通用的分析算法,这里不做限定。

[98]上述动作(b2)中,根据执行结果,在各个事件识别策略中确定问题策略。比如,根据动作(b21)中统计的数量比例,确定表示识别结果为空缺的第一识别结果数据的数量比例超过设定比例阈值的一组或多组第一识别结果数据,将该一组或多组第一识别结果数据对应的事件识别策略确定为问题策略。又如,根据动作(b22)确定的分布数据,确定对应的分布数据的分布形态不符合预设的分布形态要求的一组或多组第一识别结果数据,将该一组或多组第一识别结果数据对应的事件识别策略确定为问题策略。又如,根据动作(b23)确定的数据相关性分析结果,确定两组高度相关的第一识别结果数据,将这两组第一识别结果数据中任意一组第一识别结果数据对应的事件识别策略确定为问题策略。

[99]在一个具体的实施例中,若根据数据相关性分析结果确定存在两组第一识别结果数据的数据相关性高度相关,并且其中一组第一识别结果数据中表示识别结果为空缺的第一识别结果数据的数量比例大于预设的比例阈值,则将这两组第一识别结果数据中任意一组第一识别结果数据对应的事件识别策略确定为问题策略。

[100] 在另一个具体的实施例中,可以先确定空缺值比例要求、分布数据要求、数据相关性要求等,根据这些要求以及上述动作(b21)(b22)(b23)的执行结果,在各个事件识别策略中确定问题策略,然后在第一识别结果数据中删除问题策略对应的第一识别结果数据,并确定删除后的第一识别结果数据对应的数据推断模型。

[101] 在又一个具体的实施例中,可以先确定第一识别结果数据对应的数据推断模型,然后根据该数据推断模型的性能参数,确定对应的数据要求,数据要求包括空缺值比例要求、分布数据要求和数据相关性要求等,根据这些要求以及上述动作(b21)(b22)(b23)的执行结果,在各个事件识别策略中确定问题策略。该方式中,删除后的第一识别结果数据对应的数据推断模型,即为,删除前的第一识别结果数据对应的数据推断模型。

[102] 图3为本申请另一实施例提供的数据标签生成方法的流程示意图,如图3所示,

该流程包括以下步骤：

[103] 步骤 S302，获取目标事件的多条事件数据；

[104] 步骤 S304，通过目标事件对应的多个事件识别策略，分别对各条事件数据进行处理，得到各条事件数据的相对于风险事件的第一识别结果数据；

[105] 步骤 S306，根据第一识别结果数据，在各个事件识别策略中确定问题策略，在第一识别结果数据中删除问题策略对应的第一识别结果数据；

[106] 步骤 S308，确定删除后的第一识别结果数据对应的数据推断模型；

[107] 步骤 S310，根据删除后的第一识别结果数据对应的数据推断模型，基于删除后的第一识别结果数据进行数据推断，得到各条事件数据的相对于风险事件的第二识别结果数据；

[108] 步骤 S312，将各条事件数据的第二识别结果数据，分别确定为各条事件数据的数据标签。

[109] 图 4 为本申请另一实施例提供的数据标签生成方法的流程示意图，如图 4 所示，该流程包括以下步骤：

[110] 步骤 S402，获取目标事件的多条事件数据；

[111] 步骤 S404，通过目标事件对应的多个事件识别策略，分别对各条事件数据进行处理，得到各条事件数据的相对于风险事件的第一识别结果数据；

[112] 步骤 S406，确定第一识别结果数据对应的数据推断模型；

[113] 步骤 S408，根据第一识别结果数据和数据推断模型对应的数据要求，在各个事件识别策略中确定问题策略，在第一识别结果数据中删除问题策略对应的第一识别结果数据；

[114] 步骤 S410，根据第一识别结果数据对应的数据推断模型，基于删除后的第一识别结果数据进行数据推断，得到各条事件数据的相对于风险事件的第二识别结果数据；

[115] 步骤 S412，将各条事件数据的第二识别结果数据，分别确定为各条事件数据的数据标签。

[116] 通过图 3 图 4 中的流程，能够为各条事件数据确定准确的数据标签，从而解决人工手动打标签存在的标签准确性难以保证的问题，提高数据打标签的准确率，提高模

型的数据处理结果的准确率。尤其，通过图 4 中的流程，能够根据数据推断模型对应的数据要求在第一识别结果数据中确定不符合要求的数据进行删除，可以使得数据标签推断结果更为准确。

[117] 在资金风险识别领域中可以应用本实施例中的数据标签生成方法，通过设置目标事件为用户的交易事件，目标事件的事件数据为用户的交易事件的事件数据，可以达到为事件数据打标签的效果，其中，风险事件可以为洗钱事件、赌博事件、套现事件等。

[118] 一个具体的实施例中，选取用户的信用卡交易事件为目标事件，并设置风险事件为套现事件，从而为用户的信用卡交易数据打标签，该标签可以用于表示对应的信用卡交易事件为套现事件的概率。另一个具体的实施例中，选取用户的购物交易事件为目标事件，并设置风险事件为赌博事件，从而为用户的购物交易数据打标签，该标签可以用于表示对应的购物交易事件为赌博事件的概率。

[119] 在资金风险识别领域中通过本实施例中的方法为交易数据打标签，能够提高打标签的准确性，提高用户资金风险识别的准确性，并且由于该方法可以自动执行，不需要人工干预，因此该方法运维简单方便。

[120] 综上，本实施例中的数据标签生成方法具有以下有益效果：

[121] (1) 可以为各条事件数据确定准确的数据标签，从而解决人工手动打标签存在的标签准确性难以保证的问题，提高数据打标签的准确率，提高模型的数据处理结果的准确率；

[122] (2) 提出了一种基于弱监督学习的数据标签生成方法，针对现有的识别模型和规则准确率不高、覆盖率低的问题进行了改进，将该本质上无监督学习的问题从标签学习出发，提供了基于弱标签的解决思路和途径；

[123] (3) 该方法可以应用在资金风险识别场景中，提高用户资金风险识别的准确性；

[124] (4) 该方法可以使用多种计算机语言和软硬件实现，不受到软硬件环境和计算机语言的限制。

[125] 基于上述的数据标签生成方法，本申请实施例还提供了一种模型训练方法，图 5 为本申请一实施例提供的模型训练方法的流程示意图，如图 5 所示，该方法包括以下步骤：

[126] 步骤 S502，根据上述的数据标签生成方法，确定目标事件的事件数据的数据标

签;

[127] 这部分可以参考前面的描述, 这里不再重复。

[128] 步骤 S504, 将目标事件的事件数据作为目标事件的样本数据;

[129] 步骤 S506, 根据目标事件的样本数据和样本数据的数据标签, 训练目标事件对应的风险事件识别模型, 其中, 风险事件识别模型用于识别目标事件是否为风险事件, 或者, 用于识别目标事件为风险事件的概率。

[130] 步骤 S506 中模型可以采用已有的方法训练, 这里不做限定。本实施例中训练得到的风险事件识别模型可以为深度学习模型或卷积神经网络模型等模型。

[131] 本实施例中, 目标事件可以为用户交易事件, 目标事件的事件数据可以为用户交易事件对应的交易数据。比如, 用户通过电商平台进行多次商品购买活动, 则目标事件设定为用户的买卖交易事件, 目标事件的事件数据设定为用户的买卖交易事件对应的交易数据。本实施例中, 交易数据包括但不限于交易时间、交易金额、交易时的网络环境参数、交易所用的终端设备的设备参数、交易金额扣款途径、商家信息、收款账户信息等。

[132] 一个具体的实施例中, 选取用户的信用卡交易事件为目标事件, 并设置风险事件为套现事件, 从而为用户的信用卡交易数据打标签, 基于所打的标签训练用于识别套现事件的风险事件识别模型。另一个具体的实施例中, 选取用户的购物交易事件为目标事件, 并设置风险事件为赌博事件, 从而为用户的购物交易数据打标签, 基于所打的标签训练用于识别赌博事件的风险事件识别模型。

[133] 本实施例中, 利用上述的数据标签生成方法所生成的数据标签训练风险事件识别模型。由于在生成数据标签时通过数据推断的方式, 能够准确得到各条事件数据的识别结果数据, 因此能够准确确定事件数据的数据标签, 解决人工手动打标签存在的标签准确性难以保证的问题, 提高数据打标签的准确率, 提高模型训练的准确性, 提高模型的数据处理结果的准确率。

[134] 基于上述的模型训练方法, 本申请实施例还提供了一种事件识别方法, 图 6 为本申请一实施例提供的事件识别方法的流程示意图, 如图 6 所示, 该方法包括以下步骤:

[135] 步骤 S602, 获取待识别目标事件的事件数据。

[136] 具体地, 可以从网络上获取待识别目标事件的事件数据。

[137] 步骤 S604, 利用上述模型训练方法训练得到的风险事件识别模型, 对待识别目标事件的事件数据进行处理。

[138] 步骤 S606, 根据处理结果, 确定待识别目标事件是否为风险事件, 或者, 确定待识别目标事件为风险事件的概率。

[139] 其中, 风险事件识别模型的输出结果可以是待识别目标事件是否为风险事件, 或者是待识别目标事件为风险事件的概率。

[140] 本实施例中, 待识别目标事件可以为用户交易事件, 待识别目标事件的事件数据可以为用户交易事件对应的交易数据。比如, 用户通过电商平台进行多次商品购买活动, 则待识别目标事件设定为用户的买卖交易事件, 待识别目标事件的事件数据设定为用户的买卖交易事件对应的交易数据。本实施例中, 交易数据包括但不限于交易时间、交易金额、交易时的网络环境参数、交易所用的终端设备的设备参数、交易金额扣款途径、商家信息、收款账户信息等。

[141] 一个具体的实施例中, 选取用户的信用卡交易事件为待识别目标事件, 并设置风险事件为套现事件, 从而识别信用卡交易事件是否为套现事件。另一个具体的实施例中, 选取用户的购物交易事件为待识别目标事件, 并设置风险事件为赌博事件, 从而识别购物交易事件是否为赌博事件。

[142] 本实施例中风险事件识别模型的处理结果可以为二分类结果或分值结果。

[143] 本实施例中, 利用上述的模型训练方法训练得到的风险事件识别模型识别待识别目标事件是否为风险事件。由于在训练风险事件识别模型时, 采用上述的数据标签生成方法通过数据推断的方式准确得到各条事件数据的识别结果数据, 因此能够准确确定事件数据的数据标签, 从而提高数据打标签的准确率, 进而提高模型训练的准确性, 提高模型的数据处理结果的准确率。

[144] 基于上述的数据标签生成方法, 本申请实施例还提供了一种数据标签生成装置, 图 7 为本申请一实施例提供的一种数据标签生成装置的模块组成示意图, 如图 7 所示, 该装置包括:

[145] 第一获取模块 71, 用于获取目标事件的多条事件数据; 第一结果确定模块 72, 用于通过所述目标事件对应的多个事件识别策略, 分别对各条所述事件数据进行处理, 得到各条所述事件数据的相对于风险事件的第一识别结果数据; 第二结果确定模块 73, 用于根据所述第一识别结果数据对应的数据推断模型, 基于所述第一识别结果数据进行

数据推断，得到各条所述事件数据的相对于风险事件的第二识别结果数据；标签生成模块 74，用于根据各条所述事件数据的所述第二识别结果数据，确定各条所述事件数据的数据标签。

[146] 可选地，所述装置还包括模型确定模块，用于：在得到各条所述事件数据的相对于风险事件的第二识别结果数据之前，将每个所述事件识别策略的针对各条所述事件数据的多个所述第一识别结果数据确定为一组所述第一识别结果数据；对各组所述第一识别结果数据进行分析，以确定各组所述第一识别结果数据之间的数据相关性分析结果；分别对各组所述第一识别结果数据进行分布统计，以确定各组所述第一识别结果数据各自对应的分布数据；根据所述数据相关性分析结果和/或所述分布数据，确定所述第一识别结果数据对应的数据推断模型，所述数据推断模型包括投票模型、概率图模型和矩阵分解模型中的至少一种。

[147] 可选地，所述装置还包括数据删除模块，用于：在得到各条所述事件数据的相对于风险事件的第二识别结果数据之前，根据所述第一识别结果数据，在各个所述事件识别策略中确定问题策略，在所述第一识别结果数据中删除所述问题策略对应的第一识别结果数据；所述第二结果确定模块 73 具体用于：根据删除后的所述第一识别结果数据对应的数据推断模型，基于删除后的所述第一识别结果数据进行数据推断。

[148] 可选地，所述数据删除模块具体用于：将每个所述事件识别策略的针对各条所述事件数据的多个所述第一识别结果数据确定为一组所述第一识别结果数据；对所述第一识别结果数据分别执行以下处理中的一项或多项，根据执行结果，在各个所述事件识别策略中确定问题策略；分别统计各组所述第一识别结果数据中表示识别结果为空缺的第一识别结果数据的数量比例；分别对各组所述第一识别结果数据进行分布统计，以确定各组所述第一识别结果数据各自对应的分布数据；对各组所述第一识别结果数据进行分析，以确定各组所述第一识别结果数据之间的数据相关性分析结果。

[149] 可选地，所述数据推断模型包括投票模型、概率图模型和矩阵分解模型中的至少一种；所述第二结果确定模块 73 具体用于：通过所述数据推断模型，对各条所述事件数据对应的各个所述第一识别结果数据进行运算，得到各条所述事件数据的相对于风险事件的第二识别结果数据。

[150] 可选地，所述标签生成模块 74 具体用于：将各条所述事件数据的所述第二识别结果数据，分别确定为各条所述事件数据的数据标签。

[151] 本实施例中，首先获取目标事件的多条事件数据，并通过目标事件对应的多个事件识别策略，分别对各条事件数据进行处理，得到各条事件数据的相对于风险事件的第一识别结果数据，然后根据第一识别结果数据对应的数据推断模型，基于第一识别结果数据进行数据推断，得到各条事件数据的相对于风险事件的第二识别结果数据，最后根据各条事件数据的第二识别结果数据，确定各条事件数据的数据标签。本实施例中，通过数据推断的方式，能够准确得到各条事件数据的识别结果数据，从而准确确定事件数据的数据标签，解决人工手动打标签存在的标签准确性难以保证的问题，提高数据打标签的准确率，提高模型的数据处理结果的准确率。

[152] 基于上述的模型训练方法，本申请实施例还提供了一种模型训练装置，图 8 为本申请一实施例提供的一种模型训练装置的模块组成示意图，如图 8 所示，该装置包括：

[153] 标签确定模块 81，用于根据上述的数据标签生成装置，确定目标事件的事件数据的数据标签；

[154] 样本确定模块 82，用于将所述目标事件的事件数据作为所述目标事件的样本数据；

[155] 模型训练模块 83，用于根据所述目标事件的样本数据和所述样本数据的数据标签，训练所述目标事件对应的风险事件识别模型，其中，所述风险事件识别模型用于识别所述目标事件是否为风险事件，或者，用于识别所述目标事件为风险事件的概率。

[156] 本实施例中，利用上述的数据标签生成装置所生成的数据标签训练风险事件识别模型。由于在生成数据标签时通过数据推断的方式，能够准确得到各条事件数据的识别结果数据，因此能够准确确定事件数据的数据标签，解决人工手动打标签存在的标签准确性难以保证的问题，提高数据打标签的准确率，提高模型训练的准确性，提高模型的数据处理结果的准确率。

[157] 基于上述的事件识别方法，本申请实施例还提供了一种事件识别装置，图 9 为本申请一实施例提供的一种事件识别装置的模块组成示意图，如图 9 所示，该装置包括：

[158] 第二获取模块 91，用于获取待识别目标事件的事件数据；

[159] 数据处理模块 92，用于利用上述的模型训练装置训练得到的风险事件识别模型，对所述待识别目标事件的事件数据进行处理；

[160] 事件识别模块 93，用于根据处理结果，确定所述待识别目标事件是否为风险事件，或者，确定所述待识别目标事件为风险事件的概率。

[161] 本实施例中，利用上述的模型训练装置训练得到的风险事件识别模型识别待识别目标事件是否为风险事件。由于在训练风险事件识别模型时，采用上述的数据标签生成装置通过数据推断的方式准确得到各条事件数据的识别结果数据，因此能够准确确定事件数据的数据标签，从而提高数据打标签的准确率，进而提高模型训练的准确性，提高模型的数据处理结果的准确率。

[162] 需要说明的是，本申请实施例中的数据标签生成装置、模型训练装置和事件识别装置，分别可以实现前述的数据标签生成方法、模型训练方法和事件识别方法的实施例的各个过程，并达到相同的效果和功能，这里不再赘述。

[163] 进一步地，本申请实施例还提供了一种电子设备，图 10 为本申请一实施例提供的电子设备的结构示意图，如图 10 所示。电子设备可因配置或性能不同而产生比较大的差异，可以包括一个或一个以上的处理器 901 和存储器 902，存储器 902 中可以存储有一个或一个以上存储应用程序或数据。其中，存储器 902 可以是短暂存储或持久存储。存储在存储器 902 的应用程序可以包括一个或一个以上模块（图示未示出），每个模块可以包括对电子设备中的一系列计算机可执行指令。更进一步地，处理器 901 可以设置为与存储器 902 通信，在电子设备上执行存储器 902 中的一系列计算机可执行指令。电子设备还可以包括一个或一个以上电源 903，一个或一个以上有线或无线网络接口 904，一个或一个以上输入输出接口 905，一个或一个以上键盘 906 等。

[164] 在一个具体的实施例中，电子设备包括有存储器，以及一个或一个以上的程序，其中一个或者一个以上程序存储于存储器中，且一个或者一个以上程序可以包括一个或一个以上模块，且每个模块可以包括对电子设备中的一系列计算机可执行指令，且经配置以由一个或者一个以上处理器执行该一个或者一个以上程序包含用于进行以下计算机可执行指令：

[165] 获取目标事件的多条事件数据；

[166] 通过所述目标事件对应的多个事件识别策略，分别对各条所述事件数据进行处理，得到各条所述事件数据的相对于风险事件的第一识别结果数据；

[167] 根据所述第一识别结果数据对应的数据推断模型，基于所述第一识别结果数据进行数据推断，得到各条所述事件数据的相对于风险事件的第二识别结果数据；

[168] 根据各条所述事件数据的所述第二识别结果数据，确定各条所述事件数据的数据标签。

[169] 可选地，计算机可执行指令在被执行时，在得到各条所述事件数据的相对于风险事件的第二识别结果数据之前，还包括：将每个所述事件识别策略的针对各条所述事件数据的多个所述第一识别结果数据确定为一组所述第一识别结果数据；对各组所述第一识别结果数据进行分析，以确定各组所述第一识别结果数据之间的数据相关性分析结果；分别对各组所述第一识别结果数据进行分布统计，以确定各组所述第一识别结果数据各自对应的分布数据；根据所述数据相关性分析结果和/或所述分布数据，确定所述第一识别结果数据对应的数据推断模型，所述数据推断模型包括投票模型、概率图模型和矩阵分解模型中的至少一种。

[170] 可选地，计算机可执行指令在被执行时，在得到各条所述事件数据的相对于风险事件的第二识别结果数据之前，还包括：根据所述第一识别结果数据，在各个所述事件识别策略中确定问题策略，在所述第一识别结果数据中删除所述问题策略对应的第一识别结果数据；根据所述第一识别结果数据对应的数据推断模型，基于所述第一识别结果数据进行数据推断，包括：根据删除后的所述第一识别结果数据对应的数据推断模型，基于删除后的所述第一识别结果数据进行数据推断。

[171] 可选地，计算机可执行指令在被执行时，根据所述第一识别结果数据，在各个所述事件识别策略中确定问题策略，包括：将每个所述事件识别策略的针对各条所述事件数据的多个所述第一识别结果数据确定为一组所述第一识别结果数据；对所述第一识别结果数据分别执行以下处理中的一项或多项，根据执行结果，在各个所述事件识别策略中确定问题策略；分别统计各组所述第一识别结果数据中表示识别结果为空缺的第一识别结果数据的数量比例；分别对各组所述第一识别结果数据进行分布统计，以确定各组所述第一识别结果数据各自对应的分布数据；对各组所述第一识别结果数据进行分析，以确定各组所述第一识别结果数据之间的数据相关性分析结果。

[172] 可选地，计算机可执行指令在被执行时，所述数据推断模型包括投票模型、概率图模型和矩阵分解模型中的至少一种；根据所述第一识别结果数据对应的数据推断模型，基于所述第一识别结果数据进行数据推断，得到各条所述事件数据的相对于风险事件的第二识别结果数据，包括：通过所述数据推断模型，对各条所述事件数据对应的各个所述第一识别结果数据进行运算，得到各条所述事件数据的相对于风险事件的第二识别结果数据。

[173] 可选地，计算机可执行指令在被执行时，根据各条所述事件数据的所述第二识别结果数据，确定各条所述事件数据的数据标签，包括：将各条所述事件数据的所述第

二识别结果数据，分别确定为各条所述事件数据的数据标签。

[174] 本实施例中，首先获取目标事件的多条事件数据，并通过目标事件对应的多个事件识别策略，分别对各条事件数据进行处理，得到各条事件数据的相对于风险事件的第一识别结果数据，然后根据第一识别结果数据对应的数据推断模型，基于第一识别结果数据进行数据推断，得到各条事件数据的相对于风险事件的第二识别结果数据，最后根据各条事件数据的第二识别结果数据，确定各条事件数据的数据标签。本实施例中，通过数据推断的方式，能够准确得到各条事件数据的识别结果数据，从而准确确定事件数据的数据标签，解决人工手动打标签存在的标签准确性难以保证的问题，提高数据打标签的准确率，提高模型的数据处理结果的准确率。

[175] 在另一个具体的实施例中，电子设备包括有存储器，以及一个或一个以上的程序，其中一个或者一个以上程序存储于存储器中，且一个或者一个以上程序可以包括一个或一个以上模块，且每个模块可以包括对电子设备中的一系列计算机可执行指令，且经配置以由一个或者一个以上处理器执行该一个或者一个以上程序包含用于进行以下计算机可执行指令：

[176] 根据上述的数据标签生成方法，确定目标事件的事件数据的数据标签；

[177] 将所述目标事件的事件数据作为所述目标事件的样本数据；

[178] 根据所述目标事件的样本数据和所述样本数据的数据标签，训练所述目标事件对应的风险事件识别模型，其中，所述风险事件识别模型用于识别所述目标事件是否为风险事件，或者，用于识别所述目标事件为风险事件的概率。

[179] 本实施例中，利用上述的数据标签生成方法所生成的数据标签训练风险事件识别模型。由于在生成数据标签时通过数据推断的方式，能够准确得到各条事件数据的识别结果数据，因此能够准确确定事件数据的数据标签，解决人工手动打标签存在的标签准确性难以保证的问题，提高数据打标签的准确率，提高模型训练的准确性，提高模型的数据处理结果的准确率。

[180] 在另一个具体的实施例中，电子设备包括有存储器，以及一个或一个以上的程序，其中一个或者一个以上程序存储于存储器中，且一个或者一个以上程序可以包括一个或一个以上模块，且每个模块可以包括对电子设备中的一系列计算机可执行指令，且经配置以由一个或者一个以上处理器执行该一个或者一个以上程序包含用于进行以下计算机可执行指令：

[181] 获取待识别目标事件的事件数据;

[182] 利用上述的模型训练方法训练得到的风险事件识别模型, 对所述待识别目标事件的事件数据进行处理;

[183] 根据处理结果, 确定所述待识别目标事件是否为风险事件, 或者, 确定所述待识别目标事件为风险事件的概率。

[184] 本实施例中, 利用上述的模型训练方法训练得到的风险事件识别模型识别待识别目标事件是否为风险事件。由于在训练风险事件识别模型时, 采用上述的数据标签生成方法通过数据推断的方式准确得到各条事件数据的识别结果数据, 因此能够准确确定事件数据的数据标签, 从而提高数据打标签的准确率, 进而提高模型训练的准确性, 提高模型的数据处理结果的准确率。

[185] 需要说明的是, 本申请各个实施例中的电子设备, 分别可以实现前述的数据标签生成方法、模型训练方法和事件识别方法的实施例的各个过程, 并达到相同的效果和功能, 这里不再赘述。

[186] 进一步地, 本申请实施例还提供了一种存储介质, 用于存储计算机可执行指令, 一种具体的实施例中, 该存储介质可以为 U 盘、光盘、硬盘等, 该存储介质存储的计算机可执行指令在被处理器执行时, 能实现以下流程:

[187] 获取目标事件的多条事件数据;

[188] 通过所述目标事件对应的多个事件识别策略, 分别对各条所述事件数据进行处理, 得到各条所述事件数据的相对于风险事件的第一识别结果数据;

[189] 根据所述第一识别结果数据对应的数据推断模型, 基于所述第一识别结果数据进行数据推断, 得到各条所述事件数据的相对于风险事件的第二识别结果数据;

[190] 根据各条所述事件数据的所述第二识别结果数据, 确定各条所述事件数据的数据标签。

[191] 可选地, 该存储介质存储的计算机可执行指令在被处理器执行时, 在得到各条所述事件数据的相对于风险事件的第二识别结果数据之前, 还包括: 将每个所述事件识别策略的针对各条所述事件数据的多条所述第一识别结果数据确定为一组所述第一识别结果数据; 对各组所述第一识别结果数据进行分析, 以确定各组所述第一识别结果数据之间的数据相关性分析结果; 分别对各组所述第一识别结果数据进行分布统计, 以确

定各组所述第一识别结果数据各自对应的分布数据；根据所述数据相关性分析结果和/或所述分布数据，确定所述第一识别结果数据对应的数据推断模型，所述数据推断模型包括投票模型、概率图模型和矩阵分解模型中的至少一种。

[192] 可选地，该存储介质存储的计算机可执行指令在被处理器执行时，在得到各条所述事件数据的相对于风险事件的第二识别结果数据之前，还包括：根据所述第一识别结果数据，在各个所述事件识别策略中确定问题策略，在所述第一识别结果数据中删除所述问题策略对应的第一识别结果数据；根据所述第一识别结果数据对应的数据推断模型，基于所述第一识别结果数据进行数据推断，包括：根据删除后的所述第一识别结果数据对应的数据推断模型，基于删除后的所述第一识别结果数据进行数据推断。

[193] 可选地，该存储介质存储的计算机可执行指令在被处理器执行时，根据所述第一识别结果数据，在各个所述事件识别策略中确定问题策略，包括：将每个所述事件识别策略的针对各条所述事件数据的多个所述第一识别结果数据确定为一组所述第一识别结果数据；对所述第一识别结果数据分别执行以下处理中的一项或多项，根据执行结果，在各个所述事件识别策略中确定问题策略；分别统计各组所述第一识别结果数据中表示识别结果为空缺的第一识别结果数据的数量比例；分别对各组所述第一识别结果数据进行分布统计，以确定各组所述第一识别结果数据各自对应的分布数据；对各组所述第一识别结果数据进行分析，以确定各组所述第一识别结果数据之间的数据相关性分析结果。

[194] 可选地，该存储介质存储的计算机可执行指令在被处理器执行时，所述数据推断模型包括投票模型、概率图模型和矩阵分解模型中的至少一种；根据所述第一识别结果数据对应的数据推断模型，基于所述第一识别结果数据进行数据推断，得到各条所述事件数据的相对于风险事件的第二识别结果数据，包括：通过所述数据推断模型，对各条所述事件数据对应的各个所述第一识别结果数据进行运算，得到各条所述事件数据的相对于风险事件的第二识别结果数据。

[195] 可选地，该存储介质存储的计算机可执行指令在被处理器执行时，根据各条所述事件数据的所述第二识别结果数据，确定各条所述事件数据的数据标签，包括：将各条所述事件数据的所述第二识别结果数据，分别确定为各条所述事件数据的数据标签。

[196] 本实施例中，首先获取目标事件的多条事件数据，并通过目标事件对应的多个事件识别策略，分别对各条事件数据进行处理，得到各条事件数据的相对于风险事件的第一识别结果数据，然后根据第一识别结果数据对应的数据推断模型，基于第一识别结

果数据进行数据推断，得到各条事件数据的相对于风险事件的第二识别结果数据，最后根据各条事件数据的第二识别结果数据，确定各条事件数据的数据标签。本实施例中，通过数据推断的方式，能够准确得到各条事件数据的识别结果数据，从而准确确定事件数据的数据标签，解决人工手动打标签存在的标签准确性难以保证的问题，提高数据打标签的准确率，提高模型的数据处理结果的准确率。

[197] 在另一种具体的实施例中，该存储介质可以为 U 盘、光盘、硬盘等，该存储介质存储的计算机可执行指令在被处理器执行时，能实现以下流程：

[198] 根据上述的数据标签生成方法，确定目标事件的事件数据的数据标签；

[199] 将所述目标事件的事件数据作为所述目标事件的样本数据；

[200] 根据所述目标事件的样本数据和所述样本数据的数据标签，训练所述目标事件对应的风险事件识别模型，其中，所述风险事件识别模型用于识别所述目标事件是否为风险事件，或者，用于识别所述目标事件为风险事件的概率。

[201] 本实施例中，利用上述的数据标签生成方法所生成的数据标签训练风险事件识别模型。由于在生成数据标签时通过数据推断的方式，能够准确得到各条事件数据的识别结果数据，因此能够准确确定事件数据的数据标签，解决人工手动打标签存在的标签准确性难以保证的问题，提高数据打标签的准确率，提高模型训练的准确性，提高模型的数据处理结果的准确率。

[202] 在另一种具体的实施例中，该存储介质可以为 U 盘、光盘、硬盘等，该存储介质存储的计算机可执行指令在被处理器执行时，能实现以下流程：

[203] 获取待识别目标事件的事件数据；

[204] 利用上述的模型训练方法训练得到的风险事件识别模型，对所述待识别目标事件的事件数据进行处理；

[205] 根据处理结果，确定所述待识别目标事件是否为风险事件，或者，确定所述待识别目标事件为风险事件的概率。

[206] 本实施例中，利用上述的模型训练方法训练得到的风险事件识别模型识别待识别目标事件是否为风险事件。由于在训练风险事件识别模型时，采用上述的数据标签生成方法通过数据推断的方式准确得到各条事件数据的识别结果数据，因此能够准确确定事件数据的数据标签，从而提高数据打标签的准确率，进而提高模型训练的准确性，提

高模型的数据处理结果的准确率。

[207] 需要说明的是，本申请各个实施例中的存储介质，分别可以实现前述的数据标签生成方法、模型训练方法和事件识别方法的实施例的各个过程，并达到相同的效果和功能，这里不再赘述。

[208] 在 20 世纪 90 年代，对于一个技术的改进可以很明显地区分是硬件上的改进（例如，对二极管、晶体管、开关等电路结构的改进）还是软件上的改进（对于方法流程的改进）。然而，随着技术的发展，当今的很多方法流程的改进已经可以视为硬件电路结构的直接改进。设计人员几乎都通过将改进的方法流程编程到硬件电路中来得到相应的硬件电路结构。因此，不能说一个方法流程的改进就不能用硬件实体模块来实现。例如，可编程逻辑器件（Programmable Logic Device, PLD）（例如现场可编程门阵列（Field Programmable Gate Array, FPGA））就是这样一种集成电路，其逻辑功能由用户对器件编程来确定。由设计人员自行编程来把一个数字系统“集成”在一片 PLD 上，而不需要请芯片制造厂商来设计和制作专用的集成电路芯片。而且，如今，取代手工地制作集成电路芯片，这种编程也多半改用“逻辑编译器（logic compiler）”软件来实现，它与程序开发撰写时所用的软件编译器相类似，而要编译之前的原始代码也得用特定的编程语言来撰写，此称之为硬件描述语言（Hardware Description Language, HDL），而 HDL 也并非仅有一种，而是有许多种，如 ABEL（Advanced Boolean Expression Language）、AHDL（Altera Hardware Description Language）、Confluence、CUPL（Cornell University Programming Language）、HDCal、JHDL（Java Hardware Description Language）、Lava、Lola、MyHDL、PALASM、RHDL（Ruby Hardware Description Language）等，目前最普遍使用的是 VHDL（Very-High-Speed Integrated Circuit Hardware Description Language）与 Verilog。本领域技术人员也应该清楚，只需要将方法流程用上述几种硬件描述语言稍作逻辑编程并编程到集成电路中，就可以很容易得到实现该逻辑方法流程的硬件电路。

[209] 控制器可以按任何适当的方式实现，例如，控制器可以采取例如微处理器或处理器以及存储可由该（微）处理器执行的计算机可读程序代码（例如软件或固件）的计算机可读介质、逻辑门、开关、专用集成电路（Application Specific Integrated Circuit, ASIC）、可编程逻辑控制器和嵌入微控制器的形式，控制器的例子包括但不限于以下微控制器：ARC 625D、Atmel AT91SAM、Microchip PIC18F26K20 以及 Silicone Labs C8051F320，存储器控制器还可以被实现为存储器的控制逻辑的一部分。本领域技术人员也知道，除了以纯计算机可读程序代码方式实现控制器以外，完全可以通过将方法步

骤进行逻辑编程来使得控制器以逻辑门、开关、专用集成电路、可编程逻辑控制器和嵌入微控制器等的形式来实现相同功能。因此这种控制器可以被认为是一种硬件部件，而对其内包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至，可以将用于实现各种功能的装置视为既可以是实现方法的软件模块又可以是硬件部件内的结构。

[210] 上述实施例阐明的系统、装置、模块或单元，具体可以由计算机芯片或实体实现，或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体的，计算机例如可以为个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备的组合。

[211] 为了描述的方便，描述以上装置时以功能分为各种单元分别描述。当然，在实施本申请时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[212] 本领域内的技术人员应明白，本申请的实施例可提供为方法、系统、或计算机程序产品。因此，本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

[213] 本申请是参照根据本申请实施例的方法、设备（系统）、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[214] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[215] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在

计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[216] 在一个典型的配置中，计算设备包括一个或多个处理器（CPU）、输入/输出接口、网络接口和内存。

[217] 内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器（RAM）和/或非易失性内存等形式，如只读存储器（ROM）或闪存（flash RAM）。内存是计算机可读介质的示例。

[218] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存（PRAM）、静态随机存取存储器（SRAM）、动态随机存取存储器（DRAM）、其他类型的随机存取存储器（RAM）、只读存储器（ROM）、电可擦除可编程只读存储器（EEPROM）、快闪记忆体或其他内存技术、只读光盘只读存储器（CD-ROM）、数字多功能光盘（DVD）或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括暂存电脑可读媒体（transitory media），如调制的数据信号和载波。

[219] 还需要说明的是，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

[220] 本领域技术人员应明白，本申请的实施例可提供为方法、系统或计算机程序产品。因此，本申请可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

[221] 本申请可以在由计算机执行的计算机可执行指令的一般上下文中描述，例如程序模块。一般地，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、

对象、组件、数据结构等等。也可以在分布式计算环境中实践本申请，在这些分布式计算环境中，由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中，程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[222] 本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于系统实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[223] 以上所述仅为本申请的实施例而已，并不用于限制本申请。对于本领域技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原理之内所作的任何修改、等同替换、改进等，均应包含在本申请的权利要求范围之内。

权利要求书

1、一种数据标签生成方法，其特征在于，包括：

获取目标事件的多条事件数据；

通过所述目标事件对应的多个事件识别策略，分别对各条所述事件数据进行处理，得到各条所述事件数据的相对于风险事件的第一识别结果数据；

根据所述第一识别结果数据对应的数据推断模型，基于所述第一识别结果数据进行数据推断，得到各条所述事件数据的相对于风险事件的第二识别结果数据；

根据各条所述事件数据的所述第二识别结果数据，确定各条所述事件数据的数据标签。

2、根据权利要求1所述的方法，其特征在于，在得到各条所述事件数据的相对于风险事件的第二识别结果数据之前，所述方法还包括：

将每个所述事件识别策略的针对各条所述事件数据的多个所述第一识别结果数据确定为一组所述第一识别结果数据；

对各组所述第一识别结果数据进行分析，以确定各组所述第一识别结果数据之间的数据相关性分析结果；

分别对各组所述第一识别结果数据进行分布统计，以确定各组所述第一识别结果数据各自对应的分布数据；

根据所述数据相关性分析结果和/或所述分布数据，确定所述第一识别结果数据对应的数据推断模型，所述数据推断模型包括投票模型、概率图模型和矩阵分解模型中的至少一种。

3、根据权利要求1所述的方法，其特征在于，

在得到各条所述事件数据的相对于风险事件的第二识别结果数据之前，所述方法还包括：

根据所述第一识别结果数据，在各个所述事件识别策略中确定问题策略，在所述第一识别结果数据中删除所述问题策略对应的第一识别结果数据；

根据所述第一识别结果数据对应的数据推断模型，基于所述第一识别结果数据进行数据推断，包括：

根据删除后的所述第一识别结果数据对应的数据推断模型，基于删除后的所述第一识别结果数据进行数据推断。

4、根据权利要求3所述的方法，其特征在于，根据所述第一识别结果数据，在各个所述事件识别策略中确定问题策略，包括：

将每个所述事件识别策略的针对各条所述事件数据的多个所述第一识别结果数据确定为一组所述第一识别结果数据；

对所述第一识别结果数据分别执行以下处理中的一项或多项，根据执行结果，在各个所述事件识别策略中确定问题策略；

分别统计各组所述第一识别结果数据中表示识别结果为空缺的第一识别结果数据的数量比例；

分别对各组所述第一识别结果数据进行分布统计，以确定各组所述第一识别结果数据各自对应的分布数据；

对各组所述第一识别结果数据进行分析，以确定各组所述第一识别结果数据之间的数据相关性分析结果。

5、根据权利要求 1 至 4 任一项所述的方法，其特征在于，所述数据推断模型包括投票模型、概率图模型和矩阵分解模型中的至少一种；根据所述第一识别结果数据对应的数据推断模型，基于所述第一识别结果数据进行数据推断，得到各条所述事件数据的相对于风险事件的第二识别结果数据，包括：

通过所述数据推断模型，对各条所述事件数据对应的各个所述第一识别结果数据进行运算，得到各条所述事件数据的相对于风险事件的第二识别结果数据。

6、根据权利要求 1 至 4 任一项所述的方法，其特征在于，根据各条所述事件数据的所述第二识别结果数据，确定各条所述事件数据的数据标签，包括：

将各条所述事件数据的所述第二识别结果数据，分别确定为各条所述事件数据的数据标签。

7、一种模型训练方法，其特征在于，包括：

根据权利要求 1 至 6 中任一项所述的方法，确定目标事件的事件数据的数据标签；

将所述目标事件的事件数据作为所述目标事件的样本数据；

根据所述目标事件的样本数据和所述样本数据的数据标签，训练所述目标事件对应的风险事件识别模型，其中，所述风险事件识别模型用于识别所述目标事件是否为风险事件，或者，用于识别所述目标事件为风险事件的概率。

8、一种事件识别方法，其特征在于，包括：

获取待识别目标事件的事件数据；

利用权利要求 7 训练得到的风险事件识别模型，对所述待识别目标事件的事件数据进行处理；

根据处理结果，确定所述待识别目标事件是否为风险事件，或者，确定所述待识别

目标事件为风险事件的概率。

9、一种数据标签生成装置，其特征在于，包括：

第一获取模块，用于获取目标事件的多条事件数据；

第一结果确定模块，用于通过所述目标事件对应的多个事件识别策略，分别对各条所述事件数据进行处理，得到各条所述事件数据的相对于风险事件的第一识别结果数据；

第二结果确定模块，用于根据所述第一识别结果数据对应的数据推断模型，基于所述第一识别结果数据进行数据推断，得到各条所述事件数据的相对于风险事件的第二识别结果数据；

标签生成模块，用于根据各条所述事件数据的所述第二识别结果数据，确定各条所述事件数据的数据标签。

10、根据权利要求9所述的装置，其特征在于，所述装置还包括模型确定模块，用于：

在得到各条所述事件数据的相对于风险事件的第二识别结果数据之前，将每个所述事件识别策略的针对各条所述事件数据的多条所述第一识别结果数据确定为一组所述第一识别结果数据；

对各组所述第一识别结果数据进行分析，以确定各组所述第一识别结果数据之间的数据相关性分析结果；

分别对各组所述第一识别结果数据进行分布统计，以确定各组所述第一识别结果数据各自对应的分布数据；

根据所述数据相关性分析结果和/或所述分布数据，确定所述第一识别结果数据对应的数据推断模型，所述数据推断模型包括投票模型、概率图模型和矩阵分解模型中的至少一种。

11、根据权利要求9所述的装置，其特征在于，所述装置还包括数据删除模块，用于：

在得到各条所述事件数据的相对于风险事件的第二识别结果数据之前，根据所述第一识别结果数据，在各个所述事件识别策略中确定问题策略，在所述第一识别结果数据中删除所述问题策略对应的第一识别结果数据；

所述第二结果确定模块具体用于：

根据删除后的所述第一识别结果数据对应的数据推断模型，基于删除后的所述第一识别结果数据进行数据推断。

12、根据权利要求11所述的装置，其特征在于，所述数据删除模块具体用于：

将每个所述事件识别策略的针对各条所述事件数据的多个所述第一识别结果数据确定为一组所述第一识别结果数据；

对所述第一识别结果数据分别执行以下处理中的一项或多项，根据执行结果，在各个所述事件识别策略中确定问题策略；

分别统计各组所述第一识别结果数据中表示识别结果为空缺的第一识别结果数据的数量比例；

分别对各组所述第一识别结果数据进行分布统计，以确定各组所述第一识别结果数据各自对应的分布数据；

对各组所述第一识别结果数据进行分析，以确定各组所述第一识别结果数据之间的数据相关性分析结果。

13、根据权利要求 9 至 12 任一项所述的装置，其特征在于，所述数据推断模型包括投票模型、概率图模型和矩阵分解模型中的至少一种；所述第二结果确定模块具体用于：

通过所述数据推断模型，对各条所述事件数据对应的各个所述第一识别结果数据进行运算，得到各条所述事件数据的相对于风险事件的第二识别结果数据。

14、根据权利要求 9 至 12 任一项所述的装置，其特征在于，所述标签生成模块具体用于：

将各条所述事件数据的所述第二识别结果数据，分别确定为各条所述事件数据的数据标签。

15、一种模型训练装置，其特征在于，包括：

标签确定模块，用于根据权利要求 9 至 14 中任一项所述的装置，确定目标事件的事件数据的数据标签；

样本确定模块，用于将所述目标事件的事件数据作为所述目标事件的样本数据；

模型训练模块，用于根据所述目标事件的样本数据和所述样本数据的数据标签，训练所述目标事件对应的风险事件识别模型，其中，所述风险事件识别模型用于识别所述目标事件是否为风险事件，或者，用于识别所述目标事件为风险事件的概率。

16、一种事件识别装置，其特征在于，包括：

第二获取模块，用于获取待识别目标事件的事件数据；

数据处理模块，用于利用权利要求 15 训练得到的风险事件识别模型，对所述待识别目标事件的事件数据进行处理；

事件识别模块，用于根据处理结果，确定所述待识别目标事件是否为风险事件，或

者，确定所述待识别目标事件为风险事件的概率。

17、一种电子设备，其特征在于，包括：处理器；以及被安排成存储计算机可执行指令的存储器，所述计算机可执行指令在被执行时使所述处理器实现上述权利要求 1 至 6 任一项所述的数据标签生成方法的步骤，或者，实现上述权利要求 7 所述的模型训练方法的步骤，或者，实现上述权利要求 8 所述的事件识别方法的步骤。

18、一种存储介质，用于存储计算机可执行指令，其特征在于，所述计算机可执行指令在被执行时实现上述权利要求 1 至 6 任一项所述的数据标签生成方法的步骤，或者，实现上述权利要求 7 所述的模型训练方法的步骤，或者，实现上述权利要求 8 所述的事件识别方法的步骤。

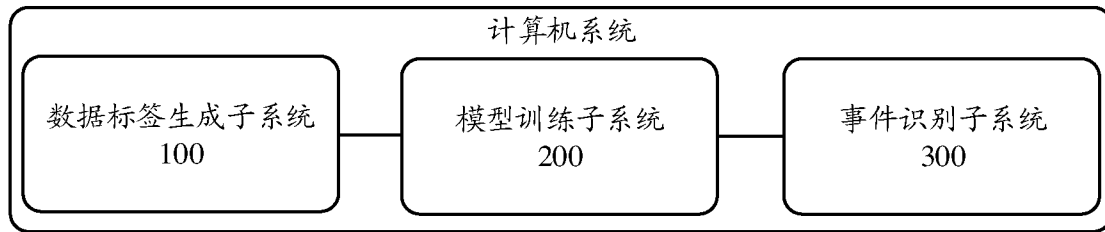


图 1

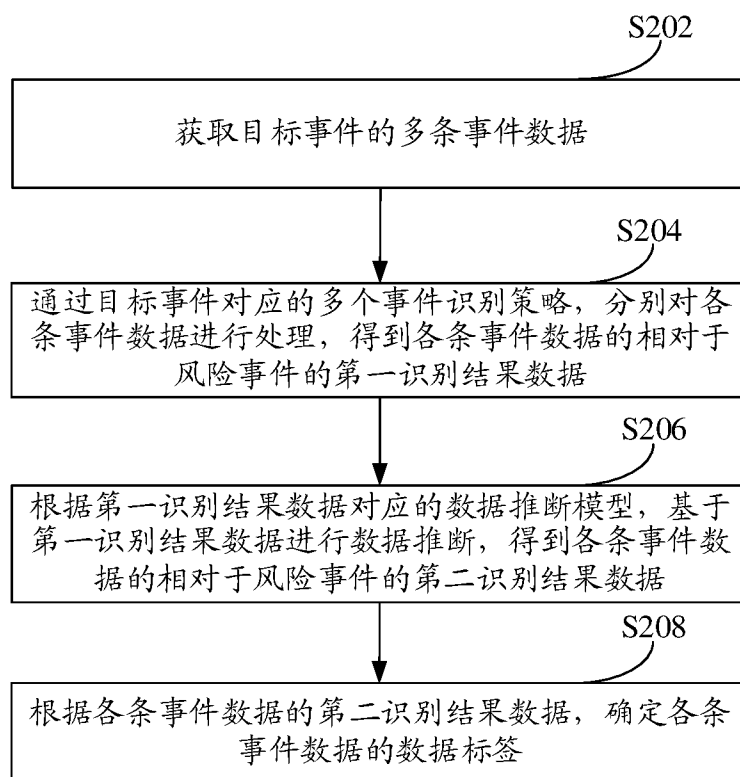


图 2

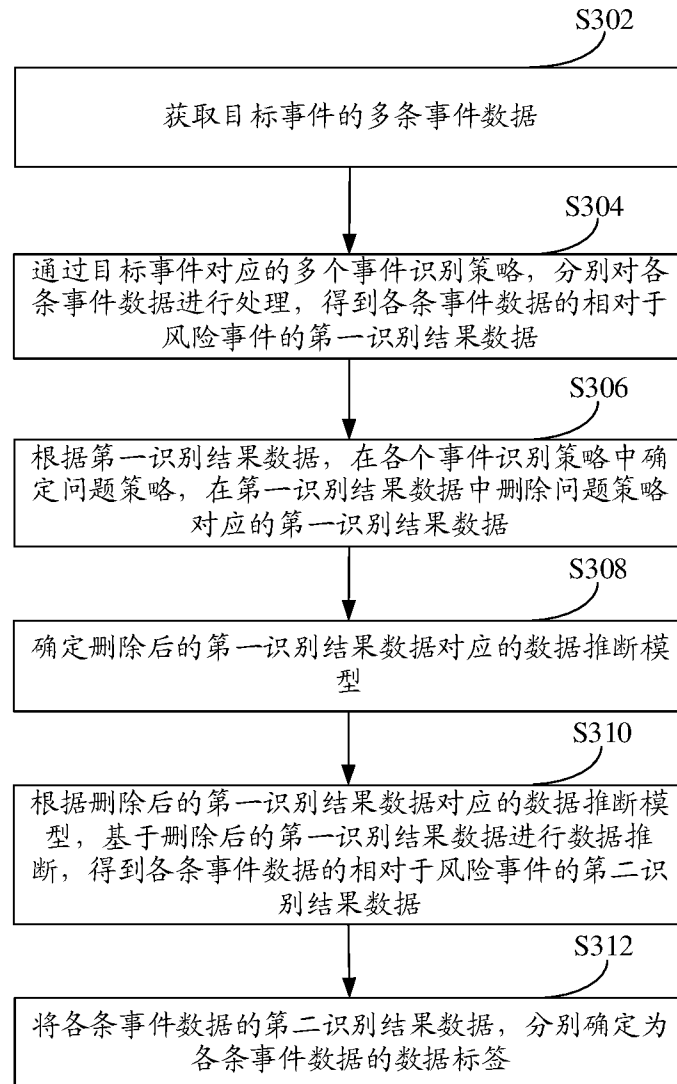


图 3

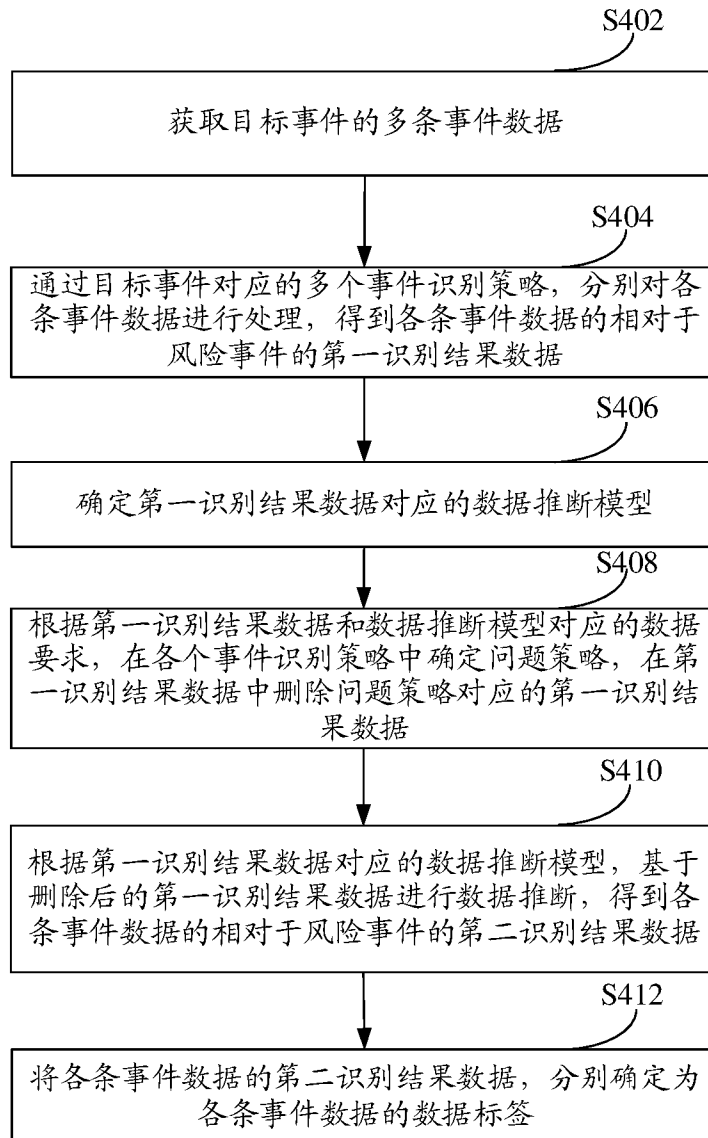


图 4

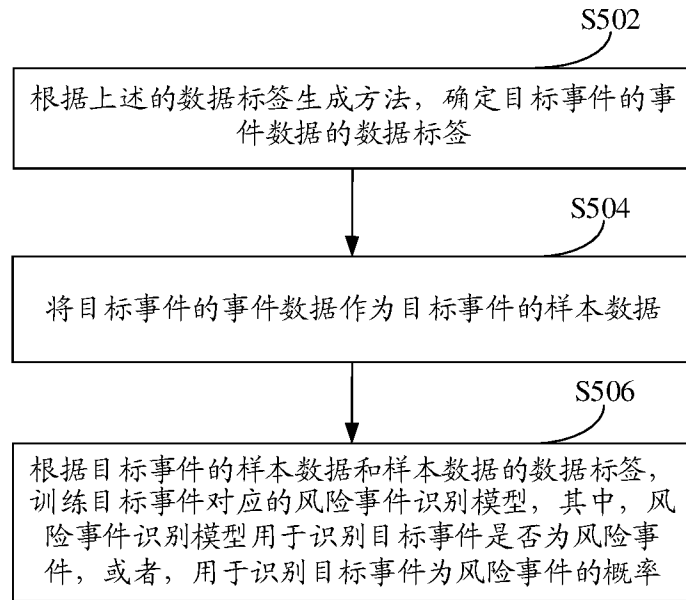


图 5

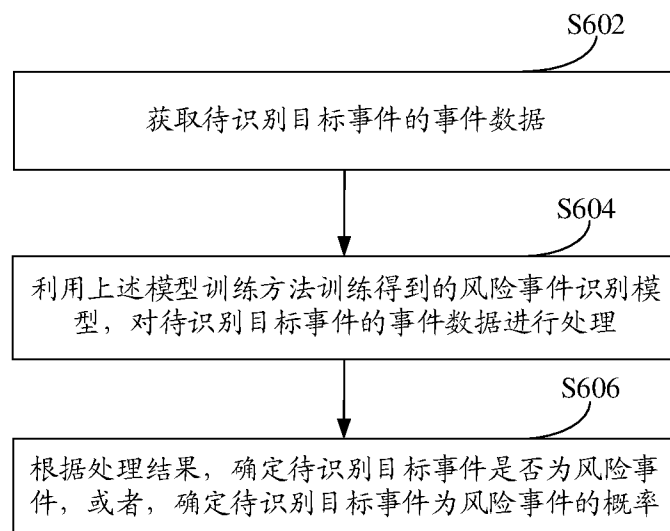


图 6

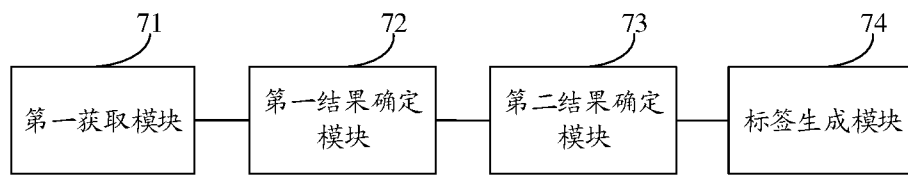


图 7

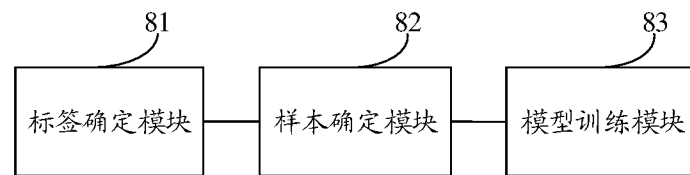


图 8

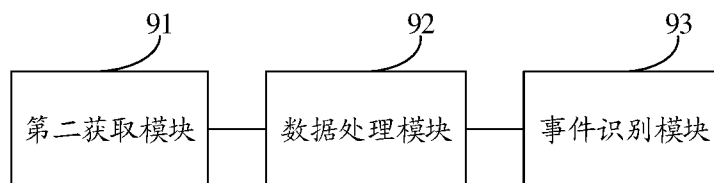


图 9

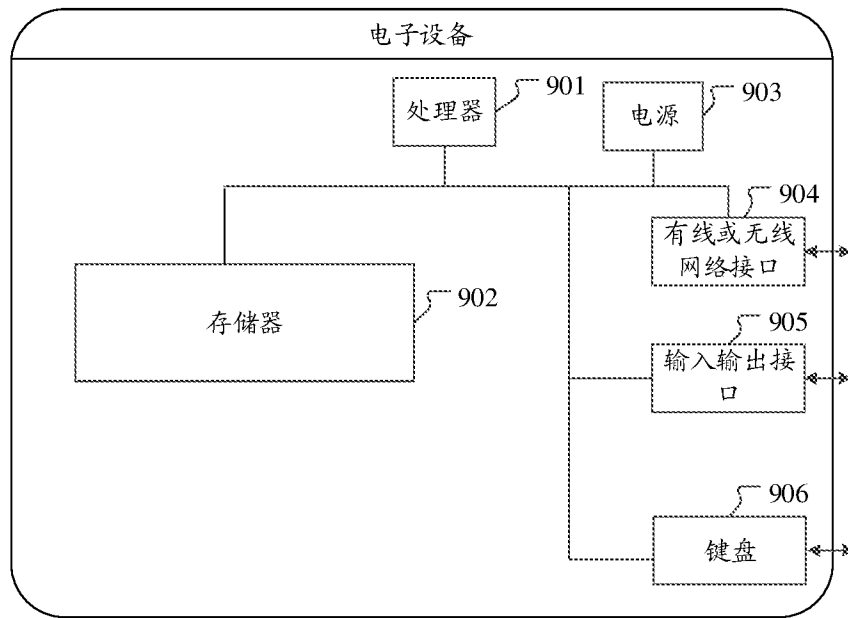


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/121902

A. CLASSIFICATION OF SUBJECT MATTER

G06N 20/00(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC: 事件; 数据; 识别; 风险; 模型; 训练; 推断; 第二; 标签; data, label, training, model, event, identificate, second, risk

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110033092 A (ALIBABA GROUP HOLDING LIMITED) 19 July 2019 (2019-07-19) claims 1-18	1-18
X	CN 108512822 A (ALIBABA GROUP HOLDING LIMITED) 07 September 2018 (2018-09-07) description, paragraphs [0009]-[0082]	1, 6, 9, 14, 17, 18
Y	CN 108512822 A (ALIBABA GROUP HOLDING LIMITED) 07 September 2018 (2018-09-07) description, paragraphs [0009]-[0082]	7, 8, 15, 16
Y	CN 109086961 A (ALIBABA GROUP HOLDING LIMITED) 25 December 2018 (2018-12-25) description, paragraphs [0005]-[0049]	7, 8, 15, 16
A	US 2016180568 A1 (BULLIVANT, DAVID PETER et al.) 23 June 2016 (2016-06-23) entire document	1-18

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

19 February 2020

Date of mailing of the international search report

06 March 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/121902

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110033092	A	19 July 2019	None			
CN	108512822	A	07 September 2018	None			
CN	109086961	A	25 December 2018	None			
US	2016180568	A1	23 June 2016	EP	3028201	A4	22 March 2017
				CN	110634175	A	31 December 2019
				US	10181213	B2	15 January 2019
				JP	2016532953	A	20 October 2016
				US	2019172242	A1	06 June 2019
				CA	2920007	A1	05 February 2015
				CN	104541306	A	22 April 2015
				WO	2015016723	A1	05 February 2015
				EP	3028201	A1	08 June 2016
				CN	104541306	B	25 October 2019

国际检索报告

国际申请号

PCT/CN2019/121902

A. 主题的分类 G06N 20/00 (2019.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类		
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06N 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNPAT, CNKI, WPI, EPDOC: 事件, 数据, 识别, 风险, 模型, 训练, 推断, 第二, 标签, data, label, training, model, event, identificate, second, risk		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 110033092 A (阿里巴巴集团控股有限公司) 2019年 7月 19日 (2019 - 07 - 19) 权利要求1-18	1-18
X	CN 108512822 A (阿里巴巴集团控股有限公司) 2018年 9月 7日 (2018 - 09 - 07) 说明书第[0009]-[0082]段	1, 6, 9, 14, 17-18
Y	CN 108512822 A (阿里巴巴集团控股有限公司) 2018年 9月 7日 (2018 - 09 - 07) 说明书第[0009]-[0082]段	7-8, 15-16
Y	CN 109086961 A (阿里巴巴集团控股有限公司) 2018年 12月 25日 (2018 - 12 - 25) 说明书第[0005]-[0049]段	7-8, 15-16
A	US 2016180568 A1 (BULLIVANT, David Peter等) 2016年 6月 23日 (2016 - 06 - 23) 全文	1-18
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期 2020年 2月 19日		国际检索报告邮寄日期 2020年 3月 6日
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 孟子山 电话号码 86-(10)-53961387

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/121902

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	110033092	A	2019年 7月 19日	无			
CN	108512822	A	2018年 9月 7日	无			
CN	109086961	A	2018年 12月 25日	无			
US	2016180568	A1	2016年 6月 23日	EP	3028201	A4	2017年 3月 22日
				CN	110634175	A	2019年 12月 31日
				US	10181213	B2	2019年 1月 15日
				JP	2016532953	A	2016年 10月 20日
				US	2019172242	A1	2019年 6月 6日
				CA	2920007	A1	2015年 2月 5日
				CN	104541306	A	2015年 4月 22日
				WO	2015016723	A1	2015年 2月 5日
				EP	3028201	A1	2016年 6月 8日
				CN	104541306	B	2019年 10月 25日