(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 9 月 10 日 (10.09.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/177377 A1

(51) 国际专利分类号:
G16H 50/80 (2018.01) *G06N 20/00* (2019.01)

(21) 国际申请号: PCT/CN2019/117290

(22) 国际申请日: 2019 年 11 月 12 日 (12.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910172323.1 2019年3月7日 (07.03.2019) CN

(71) 申请人: 平安科技（深圳）有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518033 (CN)。

(72) 发明人: 陈娴娴 (CHEN, Xianxian); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518033 (CN)。阮晓雯 (RUAN, Xiaowen); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518033 (CN)。徐亮 (XU, Liang); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518033 (CN)。

(74) 代理人: 广州华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区珠江东路6号4501房（部位：自编01-03和08-12单元）（仅限办公用途）, Guangdong 510623 (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,

(54) **Title:** MACHINE LEARNING-BASED DATA PREDICTION PROCESSING METHOD AND APPARATUS, AND COMPUTER DEVICE

(54) **发明名称:** 基于机器学习的数据预测处理方法、装置和计算机设备

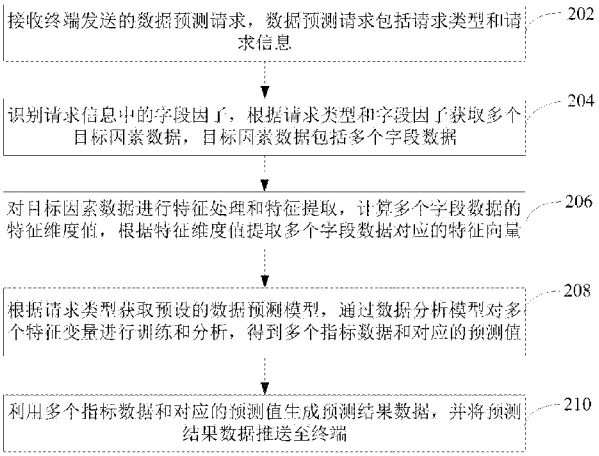

图 2

202 RECEIVE A DATA PREDICTION REQUEST SENT BY A TERMINAL, THE DATA PREDICTION REQUEST COMPRISING A REQUEST TYPE AND REQUEST INFORMATION

204 IDENTIFY A FIELD FACTOR IN THE REQUEST INFORMATION, AND ACQUIRE A PLURALITY OF PIECES OF TARGET FACTOR DATA ACCORDING TO THE REQUEST TYPE AND THE FIELD FACTOR, THE TARGET FACTOR DATA COMPRISING A PLURALITY OF PIECES OF FIELD DATA

206 PERFORM FEATURE PROCESSING AND FEATURE EXTRACTION ON THE TARGET FACTOR DATA, CALCULATE FEATURE DIMENSION VALUES OF THE PLURALITY OF PIECES OF FIELD DATA, AND EXTRACT A FEATURE VECTOR CORRESPONDING TO THE PLURALITY OF PIECES OF FIELD DATA ACCORDING TO THE FEATURE DIMENSION VALUES

208 ACQUIRE A PRESET DATA PREDICTION MODEL ACCORDING TO THE REQUEST TYPE, AND TRAIN AND ANALYZE FEATURE VARIABLES BY MEANS OF A DATA ANALYSIS MODEL TO OBTAIN A PLURALITY OF PIECES OF INDEX DATA AND CORRESPONDING PREDICTION VALUES

210 USE THE PLURALITY OF PIECES OF INDEX DATA AND THE CORRESPONDING PREDICTION VALUES TO GENERATE PREDICTION RESULT DATA, AND PUSH THE PREDICTION RESULT DATA TO THE TERMINAL

(57) **Abstract:** A machine learning-based data prediction processing method and apparatus, and a computer device, the method comprising: receiving a data prediction request sent by a terminal, the data prediction request comprising a request type and request information; identifying a field factor in the request information, and acquiring a plurality of pieces of target factor data according to the request type and the field factor, the target factor data comprising a plurality of pieces of field data; performing feature processing and feature extraction on the target factor data to obtain a feature vector corresponding to the plurality of pieces of field data; acquiring a preset data prediction model according to the request type, and training and analyzing feature variables by means of a data analysis model to obtain a plurality of pieces of index data and corresponding prediction values; and generating prediction result data according to the plurality of pieces of index data and the corresponding prediction values, and pushing the prediction result data to the terminal.

WO 2020/177377 A1

[见续页]

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：一种基于机器学习的数据预测处理方法、装置和计算机设备，所述方法包括：接收终端发送的数据预测请求，数据预测请求包括请求类型和请求信息；识别请求信息中的字段因子，根据请求类型和字段因子获取多个目标因素数据，目标因素数据包括多个字段数据；对目标因素数据进行特征处理和特征提取，得到多个字段数据对应的特征向量；根据请求类型获取预设的数据预测模型，通过数据分析模型对特征变量进行训练和分析，得到多个指标数据和对应的预测值；根据多个指标数据和对应的预测值生成的预测结果数据，并将预测结果数据推送至终端。

发明名称：基于机器学习的数据预测处理方法、装置和计算机设备

相关申请的交叉引用：

本申请要求于 2019 年 03 月 07 日提交至中国专利局，申请号为 2019101723231 ，申请名称为“基于机器学习的数据预测处理方法、装置和计算机设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及一种基于机器学习的数据预测处理方法、装置和计算机设备。

背景技术

流行性传染病是指由各种致病性微生物或病原体引起的具有传染性的疾病，如登革热、流行性病毒感冒、脑膜炎、霍乱等。由于传染病具有传播的特性，能在较短的时间内广泛蔓延，是危害人类身体健康，威胁人类生命安全的重要疾病。例如，登革热（dengue）是登革病毒经蚊媒传播引起的急性虫媒传染病，起病骤降危害大。登革热周发病人数波动较频繁，虽然每年有一定的爆发期规律，但波动周期紊乱不固定，波动幅度变换很大，因此 3-5 年长期预测的误差较难控制，预测难度非常高。

然而，随着计算机技术的迅速发展，出现了一些通过对已有的数据进行分析，从而对发病情况进行预测。但现有的方式中对发病情况进行分析的数据源较少，而影响发病情况的因素较多且有不稳定型，导致对数据进行分析预测的准确率较低。

发明内容

根据本申请公开的各种实施例，提供一种基于机器学习的数据预测处理方法、装置和计算机设备。

一种基于机器学习的数据预测处理方法包括：

接收终端发送的数据预测请求，所述数据预测请求包括请求类型和请求信息；

识别所述请求信息中的字段因子，根据所述请求类型和所述字段因子获取多个目标因素数据，所述目标因素数据包括多个字段数据；

对所述目标因素数据进行特征处理和特征提取，计算所述多个字段数据的特征维度值，

根据所述特征维度值提取所述多个字段数据对应的特征向量；

根据所述请求类型获取预设的数据预测模型，通过所述数据分析模型对所述特征变量进行训练和分析，得到多个指标数据和对应的预测值；及

利用所述多个指标数据和对应的预测值生成的预测结果数据，并将所述预测结果数据推送至所述终端。

一种基于机器学习的数据预测处理装置包括：

请求接收模块，用于接收终端发送的数据预测请求，所述数据预测请求包括请求类型和请求信息；

数据获取模块，用于识别所述请求信息中的字段因子，根据所述请求类型和所述字段因子获取多个目标因素数据，所述目标因素数据包括多个字段数据；

特征处理模块，用于对所述目标因素数据进行特征处理和特征提取，计算所述多个字段数据的特征维度值，根据所述特征维度值提取所述多个字段数据对应的特征向量；

数据预测模块，用于根据所述请求类型获取预设的数据预测模型，通过所述数据分析模型对所述特征变量进行训练和分析，得到多个指标数据和对应的预测值；及

结果数据生成模块，用于利用所述多个指标数据和对应的预测值生成的预测结果数据，并将所述预测结果数据推送至所述终端。

一种计算机设备，包括存储器和一个或多个处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述一个或多个处理器执行以下步骤：

接收终端发送的数据预测请求，所述数据预测请求包括请求类型和请求信息；

识别所述请求信息中的字段因子，根据所述请求类型和所述字段因子获取多个目标因素数据，所述目标因素数据包括多个字段数据；

对所述目标因素数据进行特征处理和特征提取，计算所述多个字段数据的特征维度值，根据所述特征维度值提取所述多个字段数据对应的特征向量；

根据所述请求类型获取预设的数据预测模型，通过所述数据分析模型对所述特征变量进行训练和分析，得到多个指标数据和对应的预测值；及

利用所述多个指标数据和对应的预测值生成的预测结果数据，将所述预测结果数据推送至所述终端。

一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行以下步骤：

接收终端发送的数据预测请求，所述数据预测请求包括请求类型和请求信息；

识别所述请求信息中的字段因子，根据所述请求类型和所述字段因子获取多个目标因素数据，所述目标因素数据包括多个字段数据；

对所述目标因素数据进行特征处理和特征提取，计算所述多个字段数据的特征维度值，根据所述特征维度值提取所述多个字段数据对应的特征向量；

根据所述请求类型获取预设的数据预测模型，通过所述数据分析模型对所述特征变量进行训练和分析，得到多个指标数据和对应的预测值；及

利用所述多个指标数据和对应的预测值生成的预测结果数据，将所述预测结果数据推送至所述终端。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征和优点将从说明书、附图以及权利要求书变得明显。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例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的前提下，还可以根据这些附图获得其它的附图。

图 1 为根据一个或多个实施例中基于机器学习的数据预测处理方法的应用场景图。

图 2 为根据一个或多个实施例中基于机器学习的数据预测处理方法的流程示意图。

图 3 为根据一个或多个实施例中训练数据预测模型步骤的流程示意图。

图 4 为根据一个或多个实施例中优化数据预测模型的步骤的流程示意图。

图 5 为根据一个或多个实施例中基于机器学习的数据预测处理装置的框图。

图 6 为根据一个或多个实施例中计算机设备的框图。

具体实施方式

为了使本申请的技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

本申请提供的基于机器学习的数据预测处理方法，可以应用于如图 1 所示的应用环境中。终端 102 通过网络与服务器 104 通过网络进行通信。终端 102 可以但不限于是各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备，服务器 104 可以用独立的服务器或者是多个服务器组成的服务器集群来实现。终端 102 可以向服务器 104 发送数据预测请

求，数据预测请求包括请求类型和请求信息，服务器 104 接收终端 102 发送的数据预测请求后，识别请求信息中的字段因子，进而根据请求类型和字段因子获取多个目标因素数据，目标因素数据包括了多个字段数据。服务器 104 则对目标因素数据进行特征处理和特征提取，得到多个字段数据对应的特征向量。服务器 104 进一步根据请求类型获取预设的数据预测模型，通过数据分析模型对多个特征向量进行分析，由此能够有效得到多个指标数据和对应的预测值。服务器 104 进而利用多个指标数据和对应的预测值生成预测结果数据，将预测结果数据推送至对应的终端 102。服务器提取出目标因素数据中的特征向量后，通过数据预测模型进行预测分析，准确有效地预测未来一段时间内目标数据的发展趋势，由此能够有效提高数据分析和预测的准确率。

在其中一个实施例中，如图 2 所示，提供了一种基于机器学习的数据预测处理方法，以该方法应用于图 1 中的服务器为例进行说明，包括以下步骤：

步骤 202，接收终端发送的数据预测请求，数据预测请求包括请求类型和请求信息。

数据预测请求可以是请求获取根据请求信息对多个目标因素数据进行分析后，得到对多个目标因素数据对应的预测结果数据。用户可以利用对应的终端向服务器发送数据预测请求，数据预测请求中包括了请求类型和请求信息，其中，请求类型可以是获取的目标预测数据的类型，例如革登热的发病预测数据；请求信息可以是用户通过终端输入的参数信息，例如时间维度参数、区域维度参数等。

步骤 204，识别请求信息中的字段因子，根据请求类型和字段因子获取多个目标因素数据，目标因素数据包括多个字段数据。

服务器接收到终端发送的数据预测请求后，对数据预测请求中的请求信息进行分析。具体地，服务器获取请求信息中的参数信息，对参数信息进行解析，识别参数信息中的字段因子，例如，字段因子可以包括时间因子、区域因子、天气因子等。服务器进而根据请求类型和识别出的字段因子从本地数据库中获取对应的多个目标因素数据，服务器还可以从第三方数据库中获取对应的多个目标因素数据，目标因素数据中包括了多个字段数据。目标因素数据可以是与字段因子相对应的数据，例如当请求类型为革登热发病趋势预测时，则目标因素数据可以包括天气数据、环境数据以及疾病类型对应的历史发病数据等，例如获取未来一周的天气数据，包括周平均温度、周最大气压、周降水量等。

步骤 206，对目标因素数据进行特征处理和特征提取，计算多个字段数据的特征维度值，根据特征维度值提取多个字段数据对应的特征向量。

服务器获取多个目标因素数据后，则对目标因素数据进行处理。具体地，服务器首先对

目标因素数据对应的多个字段数据进行向量化，将多个目标因素数据按照时间序列转换为对应的特征向量，并将多个特征向量转换为多个特征变量和对应的维度特征值，维度特征值可以表示为特征变量所属的特征维度，由此得到多个字段数据对应的特征向量。服务器进一步对特征向量进行数据清洗和衍生处理，得到处理后的多个特征向量。

例如，服务器可以通过均值填充、定制填充和书模型填充等方式对多个特征向量进行预处理，并通过数据均值、方差和标准差等对多个特征向量处理，提取出多个特征向量和维度特征值。服务器可以进一步对特征向量进行缺失值填充、异常值抽取更替等。例如，得到多个维度的特征向量后，这些数据背后可能还有隐藏的信息，因此还可以采用一些特征工程的算法，如统计学相关的特征衍生，通过对每一维度的特征向量作滞后分析，将每一维度 n 维的滞后数据作方差、标准差、均值等的衍生，进而得到目标因素数据对于的衍生特征变量。

服务器得到处理后的特征向量后，进一步对目标因素数据进行特征提取，得到对应的多个字段数据对应的特征向量，并根据预设的目标函数计算特征向量的多个特征维度值。具体地，根据预设的特征降维算法对多个特征向量进行特征提取，特征维度值计算多个特征向量之间的相似度，提取出相似度达到阈值的特征向量。例如，可以利用奇异值分解、主成分分析等算法对特征变量整体进行降维，从而能够有效地对目标因素数据进行特征提取，提取出有价值的特征向量和对应的维度特征值。

步骤 208，根据请求类型获取预设的数据预测模型，通过数据分析模型对多个特征变量进行训练和分析，得到多个指标数据和对应的预测值。

步骤 210，利用多个指标数据和对应的预测值生成预测结果数据，将预测结果数据推送至终端。

服务器提取出目标因素数据中对应的特征向量后，则根据请求类型获取预设的数据预测模型。数据预测模型可以是基于机器学习的神经网络模型，数据预测模型中可以包括预设的输入层、LSTM 层、Dropout 层和输出层。服务器则将多个特征向量和对应的维度特征值输入至数据预测模型的输入层中，通过数据预测模型对输入的数据进行训练。具体地，通过数据预测模型中的预设函数计算多个特征向量对应的权重，并计算多个特征向量与目标特征数据之间的相关性。服务器进而利用预设算法根据多个特征向量与目标特征数据之间的相关性进行回归分析，通过数据预测模型的输出层输出多个预设的指标数据和对应的预测值，进而根据多个指标数据和对应的预测值生成与数据预测请求相对应的预测结果数据。

例如，当终端向服务器发送的数据预测请求是基于革登热发病趋势的预测请求时，服务器接收到数据预测请求后，根据请求类型和请求信息获取对应时间段和区域的天气数据和环

境数据等，例如，天气数据可以包括周平均温度、周最大气压、周降水量等。服务器则对获取的目标因素数据进行特征处理和特征提取，提取出目标因素数据中对应的特征向量。服务器进而将提取出的目标因素数据对应的特征向量输入至数据预测模型中进行学习和训练，并进行回归分析，分析出在预设时间段内革登热的发病预测趋势。

服务器进而将预测结果数据发送至对应终端。进一步地，服务器还可以将预测结果数据按照预设方式生成对应的视图预测数据。具体地，服务器可以根据数据预测请求中的请求类型获取对应的集成函数，并根据预测结果数据中的多个预设时序参数和对应的预测值通过集成函数集成对应的视图预测数据，服务器进而将视图预测数据推送至终端。通过提取出目标因素数据中的特征向量后，通过数据预测模型进行分析，准确有效地预测未来一段时间内的目标数据发展趋势，由此能够有效提高数据分析和预测的准确率。

上述基于机器学习的数据预测方法中，终端可以向服务器发送数据预测请求，数据预测请求包括请求类型和请求信息，服务器接收终端发送的数据预测请求后，识别请求信息中的字段因子，进而根据请求类型和字段因子获取多个目标因素数据，目标因素数据包括了多个字段数据。服务器则对目标因素数据进行特征处理和特征提取，计算多个字段数据的特征维度值，根据特征维度值提取多个字段数据对应的特征向量。服务器进一步根据请求类型获取预设的数据预测模型，通过数据分析模型对多个特征向量进行分析，由此能够有效得到多个指标数据和对应的预测值。服务器进而根据多个指标数据和对应的预测值生成预测结果数据，将预测结果数据推送至对应的终端。服务器提取出目标因素数据中的特征向量后，通过数据预测模型进行分析，准确有效地预测未来一段时间内目标数据的发展趋势，由此能够有效提高数据分析和预测的准确率。

在其中一个实施例中，对目标因素数据进行特征处理和特征提取，计算多个字段数据的特征维度值，根据特征维度值提取多个字段数据对应的特征向量的步骤，包括：对目标因素数据对应的多个字段数据进行向量化，得到多个字段数据对应的特征向量；对特征向量进行衍生处理，得到处理后的多个特征向量；根据预设目标函数计算特征向量的多个特征维度值；根据预设距离算法和特征维度值计算多个特征向量之间的相似度；提取出相似度达到预设阈值的特征向量。

用户可以利用对应的终端向服务器发送数据预测请求，数据预测请求中包括了请求类型和请求信息。服务器接收到终端发送的数据预测请求后，对数据预测请求中的请求信息进行分析。具体地，服务器获取请求信息中的参数信息，对参数信息进行解析，识别参数信息中的字段因子。服务器进而根据请求类型和识别出的字段因子从本地数据库中获取对应的多个

目标因素数据，服务器还可以从第三方数据库中获取对应的多个目标因素数据，目标因素数据中包括了多个字段数据。

服务器获取多个目标因素数据后，则对目标因素数据进行处理。具体地，服务器首先对目标因素数据对应的多个字段数据进行向量化，将多个目标因素数据按照时间序列转换为对应的特征向量。服务器进而根据预设的目标函数计算多个特征向量对应的特征维度值，特征维度值可以表示为特征向量所属的特征维度，由此得到多个字段数据对应的特征向量。服务器进一步对特征向量进行数据清洗和衍生处理，得到处理后的多个特征向量。

例如，服务器可以通过均值填充、定制填充和书模型填充等方式对多个特征向量进行预处理，并通过数据均值、方差和标准差等对多个特征向量处理，提取出多个特征向量和计算出对应的特征维度值。服务器可以进一步对特征向量进行缺失值填充、异常值抽取更替等。例如，得到多个维度的特征向量后，这些数据背后可能还有隐藏的信息，因此还可以采用一些特征工程的算法，如统计学相关的特征衍生，通过对每一维度的特征向量作滞后分析，将每一维度 n 维的滞后数据作方差、标准差、均值等等的衍生，进而得到目标因素数据对于的衍生特征向量。

服务器得到处理后的特征向量后，进一步对目标因素数据进行特征提取，具体地，根据预设的距离算法和特征维度值计算多个特征向量之间的相似度，提取出相似度达到阈值的特征向量。服务器还可以进一步利用奇异值分解、主成分分析等算法对特征向量整体进行降维，从而能够有效地对目标因素数据进行特征提取，提取出有价值的特征向量和对应的特征维度值。

在其中一个实施例中，如图 3 所示，获取预设的数据预测模型之前，还包括构建和训练数据预测模型的步骤，该步骤具体包括以下内容：

步骤 302，获取多个历史因素数据，根据多个历史因素数据生成对应的训练集数据和验证集数据。

服务器在获取预设的数据预测模型之前，还需要构建和训练出数据预测模型。具体地，服务器可以预先从本地数据库或第三方数据库中获取大量的历史发病数据、天气数据以及环境数据等多个历史因素数据，例如服务器可以获取 2000-2018 年的登革热的周发病人数、每周的天气数据，包括周平均温度、周最大气压、周降水量等历史数据。服务器则将历史因素数据生成训练集数据和验证集数据。训练集数据可以是经过人工标注后的数据。

步骤 304，对练集数据中的多个历史因素数据进行聚类分析，得到聚类结果。

步骤 306，对聚类结果进行特征提取，提取出多个特征变量。

步骤 308，获取预设的神经网络模型，通过神经网络模型对训练集数据进行训练，得到多个特征变量对应的特征维度值和权重，根据多个特征变量对应的特征维度值和权重构建初始数据预测模型。

步骤 310，利用验证集数据对所述数据预测模型进行进一步训练和验证，直到验证集数据中满足预设阈值的数量达到预设比值时，停止训练，得到训练完成的数据预测模型。

服务器首先对训练集数据中的历史因素数据进行数据清洗和数据预处理，具体地，服务器对历史因素数据对应的多个字段数据进行向量化，得到多个字段数据对应的特征向量，并将特征向量转换为对应的特征变量。服务器进一步对特征变量进行衍生处理，得到处理后的多个特征变量。如对特征变量进行缺失值填充、异常值抽取更替等。

服务器进而采用预设的聚类算法对训练集数据对应的特征变量进行聚类分析。例如，预设的聚类算法可以为 k-means（k-均值算法）聚类的方法。服务器通过对特征变量进行多次聚类后得到多个聚类结果。服务器并根据预设算法计算多个特征变量之间的相似度，提取出相似度达到预设阈值的特征变量。

例如，服务器可以对多个聚类结果内的特征变量分别进行组合，得到多个组合特征变量。获取目标变量，利用目标变量对多个组合特征变量进行相关性检验。检验通过时，对组合特征变量添加交互标签。利用添加交互标签后的组合特征变量解析对应的特征变量。添加交互标签后的组合特征变量可以为达到预设阈值的特征变量，服务器则提取出达到预设阈值特征变量。通过对特征变量进行特征处理和特征提取，能够有效地提取出有价值的特征变量。

服务器则获取预设的神经网络模型。例如，神经网络模型可以包括预设的输入层、多个 LSTM 层、dropout 层和输出层。神经网络模型中包括多个网络节点，其中每一层网络节点的舍弃率可以为 0.2。神经网络模型的 LSTM 层包括激活函数，通过 LSTM 层输出的完全连接人工神经网络也包括对应的激活函数。神经网络模型中还包括确定误差的计算方式，例如可以采用均方误差算法；还包括确定权重参数的迭代更新方式，例如可以采用 RMSprop 算法。神经网络模型中还包括回归模型，通过对数据进行预测分析后，进而可以得到连续的数据预测值。进一步地，神经网络模型中还可以包括一层普通的神经网络层，用于输出结果的降维。

服务器获取预设的神经网络模型后，进一步将提取出的训练集数据对应的特征向量输入至神经网络模型中进行学习和训练。服务器通过对训练集中的大量发病历史数据进行训练后，可以得到得到多个特征变量对应的特征维度值和权重，进而根据多个特征变量对应的特征维度值和权重构建初始数据预测模型。

服务器得到初始数据预测模型后，则获取验证集数据，通过验证集数据中的大量历史因

素数据对构建的初始数据预测模型进行训练和验证。直到验证集数据中满足预设阈值的数量达到预设比值时，则停止训练，进而得到训练完成的数据预测模型。通过对大量的历史因素数据进行训练和学习，进而能够有效构建出预测准确率较高的数据预测模型，从而有效提高了数据预测的准确率。

在其中一个实施例中，通过数据预测模型对多个特征向量进行分析，得到对应的预测结果数据的步骤，包括：将多个特征向量作为数据预测模型的输入，根据数据预测模型中的预设函数计算多个特征向量对应的权重；根据权重计算多个特征向量与目标特征数据之间的相关性；根据权重和相关性进行回归分析，通过数据预测模型中的输出层输出多个指标数据和对应的预测值。

用户可以利用对应的终端向服务器发送数据预测请求，数据预测请求中包括了请求类型和请求信息。服务器接收到终端发送的数据预测请求后，对数据预测请求中的请求信息进行分析。具体地，服务器获取请求信息中的参数信息，对参数信息进行解析，识别参数信息中的字段因子。服务器进而根据请求类型和识别出的字段因子从本地数据库中获取对应的多个目标因素数据，服务器还可以从第三方数据库中获取对应的多个目标因素数据，目标因素数据中包括了多个字段数据。

服务器获取多个目标因素数据后，对目标因素数据进行特征处理和特征提取，得到多个字段数据对应的特征向量。服务器则进一步获取预设的数据预测模型，将多个字段数据对应的特征向量作为数据预测模型的输入，进而通过数据预测模型对多个特征向量进行学习和训练，根据数据预测模型中的预设函数计算出多个特征向量对应的权重，并根据权重计算出多个特征向量与目标特征数据的相关性。数据预测模型中还包括回归分析模型，服务器进而通过数据预测模型中的预设算法根据多个特征向量的权重以及多个特征向量与多个预设目标特征数据之间的相关性进行回归分析，由此可以分析得到多个指标数据和对应的预测值。进而通过数据预测模型中的输出层输出多个指标数据和对应的预测值。服务器则利用多个指标数据和对应的预测值生成与数据预测请求对应的预测结果数据。通过利用预先训练的数据预测模型对目标因素数据进行预测分析，能够有效地提高数据预测的准确率。

在其中一个实施例中，如图 4 所示，该方法还包括对数据预测模型进行优化的步骤，该步骤具体包括以下内容：

步骤 402，根据预设频率从预设数据库中获取多个历史因素数据。

步骤 404，对多个历史因素数据进行聚类分析，得到分析结果。

步骤 406，根据所述分析结果进行特征选择，得到多个特征变量。

步骤408，根据预设的算法计算多个特征变量的权重。

步骤410，根据多个特征变量和对应的权重对所述数据预测模型进行优化调整。

服务器训练得到数据预测模型后，还可以根据预设的频率对数据预测模型进行调参优化。具体地，服务器可以根据预设频率从数据库中获取大量的历史因素数据，例如预设频率可以为一周、两周、一个月、三个月等，服务器则可以获取过去一周、两周、一个月、三个月内的历史因素数据，历史因素数据可以包括预设区域的发病数据、天气数据以及环境数据等，例如，天气数据可以包括周平均温度、周最大气压、周降水量等。

服务器首先获取的大量历史因素数据进行数据清洗和数据预处理，具体地，服务器对历史因素数据对应的多个字段数据进行向量化，得到多个字段数据对应的特征变量，并对特征变量进行衍生处理，得到处理后的多个特征变量。如对特征变量进行缺失值填充、异常值抽取更替等。

服务器进而采用预设的聚类算法对训练集数据对应的特征变量进行聚类分析。例如，预设的聚类算法可以为k-means（k-均值算法）聚类的方法。服务器通过对特征变量进行多次聚类后得到多个聚类结果。服务器并根据预设算法计算多个特征变量之间的相似度，提取出相似度达到预设阈值的特征变量。

例如，服务器可以对多个聚类结果内的特征变量分别进行组合，得到多个组合特征变量。获取目标变量，利用目标变量对多个组合特征变量进行相关性检验。检验通过时，对组合特征变量添加交互标签。利用添加交互标签后的组合特征变量解析对应的特征变量。添加交互标签后的组合特征变量可以为达到预设阈值的特征变量，服务器则提取出达到预设阈值特征变量。通过对特征变量进行特征处理和特征提取，能够有效地提取出有价值的特征变量。

服务器进一步根据预设的算法计算多个特征变量的权重，进而根据多个特征变量和对应的权重对数据预测模型进行优化调整。具体地，服务器可以根据多个特征变量和对应的权重对数据预测模型中的模型参数进行调整，由此能够有效地对数据预测模型进行调参优化。

在其中一个实施例中，该方法还包括：根据请求类型获取预设的集成函数；根据预测结果数据中的多个预设时序参数和对应的预测值通过集成函数集成对应的视图预测数据；对视图预测数据添加事件类型标识和对应的接口调用参数，将视图预测数据发送至终端。

终端可以向服务器发送数据预测请求，数据预测请求包括请求类型和请求信息，服务器接收终端发送的数据预测请求后，识别请求信息中的字段因子，进而根据请求类型和字段因子获取多个目标因素数据，目标因素数据包括了多个字段数据。服务器则对目标因素数据进行特征处理和特征提取，得到多个字段数据对应的特征变量。服务器进一步根据请求类型获

取预设的数据预测模型，通过数据分析模型对多个特征变量进行分析，得到多个指标数据和对应的预测值，服务器进而根据多个指标数据和对应的预测值生成预测结果数据，由此准确有效地预测未来一段时间内目标数据的发展趋势，进而能够有效提高数据分析和预测的准确率。

服务器生成与数据预测请求对应的预测结果数据后，还可以进一步将预测结果数据转换为对应的视图预测数据。具体地，请求类型中包括了获取的目标预测结果数据的类型，预测结果数据中包括了多个指标数据和对应的预测值。例如，可以包括基于时间维度的发病概率、发病分布情况等参数以及对应的预测值。如时间维度可以是每 3 小时、每 12 小时为时间单位每天或者每周为时间单位。服务器可以根据请求类型获取预设的集成函数，例如 python 可视化函数，将多个指标数据和对应的预测值集成对应的视图数据，例如可以利用直方图可视化函数、分布密度、热度图等可视化函数嵌入集成对应的视图数据，通过嵌套函数能够绘制出对应的可视化图像。

服务器根据预测结果数据中的多个指标数据和对应的预测值通过集成函数集成对应的视图预测数据后，进一步对视图资源数据添加事件类型标识和对应的接口调用参数，并集成对应的类进行存储。以利于服务器或终端对生成的视图预测数据进行调用，由此使得服务器或终端再次获取相关联的时序数据或视图数据时，可以直接根据事件类型标识和对应的接口调用参数调用挖掘分析出的视图资源数据，进而提高了对时序数据的分析效率和利用价值。

服务器生成对应的视图预测数据后，则将视图预测数据推送至对应的终端。服务器对目标因素数据进行分析和预测，并得到对应的预测结果数据后，通过将预测结果数据转换为对应的视图数据，由此能够有效地对预测结果数据进行可视化，使得用户能够有效地获取和查看预测结果数据。

应该理解的是，虽然图 2-4 的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，图 2-4 中的至少一部分步骤可以包括多个子步骤或者多个阶段，这些子步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些子步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

在其中一个实施例中，如图 5 所示，提供了一种基于机器学习的数据预测处理装置，包括：请求接收模块 502、数据获取模块 504、特征处理模块 506、数据预测模块 508 和结果数据生成模块 510，其中：

请求接收模块 502，用于接收终端发送的数据预测请求，数据预测请求包括请求类型和请求信息；

数据获取模块 504，用于识别请求信息中的字段因子，根据请求类型和字段因子获取多个目标因素数据，目标因素数据包括多个字段数据；

特征处理模块 506，用于对目标因素数据进行特征处理和特征提取，计算多个字段数据的特征维度值，根据特征维度值提取多个字段数据对应的特征向量；

数据预测模块 508，用于根据请求类型获取预设的数据预测模型，通过数据分析模型对特征变量进行训练和分析，得到多个指标数据和对应的预测值；

结果数据生成模块 510，用于利用多个指标数据和对应的预测值生成的预测结果数据，并将预测结果数据推送至终端。

在其中一个实施例中，特征处理模块 506 还用于对目标因素数据对应的多个字段数据进行向量化，得到多个字段数据对应的特征向量；对特征变量进行衍生处理，得到处理后的多个特征向量；根据预设目标函数计算特征向量的多个特征维度值；根据预设距离算法和特征维度值计算多个特征向量之间的相似度；提取出相似度达到预设阈值的特征向量。

在其中一个实施例中，该装置还包括模型训练模块，用于获取多个历史因素数据，根据所述多个历史因素数据生成对应的训练集数据和验证集数据；对训练集数据中的多个历史因素数据进行聚类分析，得到聚类结果；对聚类结果进行特征提取，提取出多个特征变量；获取预设的神经网络模型，通过神经网络模型对训练集数据进行训练，得到多个特征变量对应的特征维度值和权重，根据多个特征变量对应的特征维度值和权重构建初始数据预测模型；利用验证集数据对数据预测模型进行进一步训练和验证，直到验证集数据中满足预设阈值的数量达到预设比值时，停止训练，得到训练完成的数据预测模型。

在其中一个实施例中，数据预测模块 508 还用于将多个特征向量作为数据预测模型的输入，根据数据预测模型中的预设函数计算多个特征向量对应的权重；根据权重计算多个特征向量与目标特征数据之间的相关性；根据特征向量的权重和相关性进行回归分析，并通过数据预测模型中的输出层输出多个指标数据和对应的预测值。

在其中一个实施例中，该装置还包括模型优化模块，用于根据预设频率从预设数据库中获取多个历史因素数据；对多个历史因素数据进行聚类分析，得到分析结果；根据分析结果进行特征选择，得到多个特征变量；根据预设的算法计算多个特征变量的权重；根据多个特征变量和对应的权重对数据预测模型进行优化调整。

在其中一个实施例中，该装置还包括视图集成模块，用于根据请求类型获取预设的集成

函数；根据预测结果数据中的多个预设时序参数和对应的预测值通过集成函数集成对应的视图预测数据；对视图预测数据添加事件类型标识和对应的接口调用参数，并将视图预测数据发送至终端。

关于基于机器学习的数据预测处理装置的具体限定可以参见上文中对于基于机器学习的数据预测处理方法的限定，在此不再赘述。上述基于机器学习的数据预测处理装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

在一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图 6 所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储目标因素数据、历史因素数据、预测结果数据以及试图结果数据等。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现本申请任意一个实施例中提供的基于机器学习的数据预测处理方法的步骤。

本领域技术人员可以理解，图 6 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

一种计算机设备，包括存储器和一个或多个处理器，存储器中储存有计算机可读指令，计算机可读指令被处理器执行时，使得一个或多个处理器执行以下步骤：

接收终端发送的数据预测请求，数据预测请求包括请求类型和请求信息；

识别请求信息中的字段因子，根据请求类型和字段因子获取多个目标因素数据，目标因素数据包括多个字段数据；

对目标因素数据进行特征处理和特征提取，计算多个字段数据的特征维度值，根据特征维度值提取多个字段数据对应的特征向量；

根据请求类型获取预设的数据预测模型，通过数据分析模型对特征变量进行训练和分析，得到多个指标数据和对应的预测值；及

利用多个指标数据和对应的预测值生成的预测结果数据，将预测结果数据推送至终端。

一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行以下步骤：

接收终端发送的数据预测请求，数据预测请求包括请求类型和请求信息；

识别请求信息中的字段因子，根据请求类型和字段因子获取多个目标因素数据，目标因素数据包括多个字段数据；

对目标因素数据进行特征处理和特征提取，计算多个字段数据的特征维度值，根据特征维度值提取多个字段数据对应的特征向量；

根据请求类型获取预设的数据预测模型，通过数据分析模型对特征变量进行训练和分析，得到多个指标数据和对应的预测值；及

利用多个指标数据和对应的预测值生成的预测结果数据，将预测结果数据推送至终端。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性计算机可读取存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器（ROM）、可编程ROM（PROM）、电可编程ROM（EPROM）、电可擦除可编程ROM（EEPROM）或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM以多种形式可得，诸如静态RAM（SRAM）、动态RAM（DRAM）、同步DRAM（SDRAM）、双数据率SDRAM（DDRSDRAM）、增强型SDRAM（ESDRAM）、同步链路（Synchlink）DRAM（SLDRAM）、存储器总线（Rambus）直接RAM（RDRAM）、直接存储器总线动态RAM（DRDRAM）、以及存储器总线动态RAM（RDRAM）等。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求书

1、一种基于机器学习的数据预测处理方法，所述方法包括：

接收终端发送的数据预测请求，所述数据预测请求包括请求类型和请求信息；

识别所述请求信息中的字段因子，根据所述请求类型和所述字段因子获取多个目标因素数据，所述目标因素数据包括多个字段数据；

对所述目标因素数据进行特征处理和特征提取，计算所述多个字段数据的特征维度值，根据所述特征维度值提取所述多个字段数据对应的特征向量；

根据所述请求类型获取预设的数据预测模型，通过所述数据分析模型对所述特征变量进行训练和分析，得到多个指标数据和对应的预测值；及

利用所述多个指标数据和对应的预测值生成的预测结果数据，将所述预测结果数据推送至所述终端。

2、根据权利要求 1 所述的方法，其特征在于，所述对所述目标因素数据进行特征处理和特征提取，计算所述多个字段数据的特征维度值，根据所述特征维度值提取所述多个字段数据对应的特征向量的步骤，包括：

对所述目标因素数据对应的多个字段数据进行向量化，得到多个字段数据对应的特征向量；

对所述特征变量进行衍生处理，得到处理后的多个特征向量；

根据预设目标函数计算所述特征向量的多个特征维度值；

根据预设距离算法和所述特征维度值计算多个特征向量之间的相似度；及

提取出所述相似度达到预设阈值的特征向量。

3、根据权利要求 1 所述的方法，其特征在于，所述获取预设的数据预测模型的步骤之前，还包括：

获取多个历史因素数据，根据所述多个历史因素数据生成对应的训练集数据和验证集数据；

对所述训练集数据中的多个历史因素数据进行聚类分析，得到聚类结果；

对所述聚类结果进行特征提取，提取出多个特征变量；

获取预设的神经网络模型，通过所述神经网络模型对所述训练集数据进行训练，得到多个特征变量对应的特征维度值和权重，根据多个特征变量对应的特征维度值和权重构建初始数据预测模型；及

利用所述验证集数据对所述数据预测模型进行进一步训练和验证，直到所述验证集数据

中满足预设阈值的数量达到预设比值时，停止训练，得到训练完成的数据预测模型。

4、根据权利要求 1 所述的方法，其特征在于，所述通过所述数据分析模型对所述特征变量进行训练和分析，得到多个指标数据和对应的预测值的步骤，包括：

将所述多个特征向量作为所述数据预测模型的输入，根据所述数据预测模型中的预设函数计算多个特征向量对应的权重；

根据所述权重计算多个特征向量与目标特征数据之间的相关性；及

根据所述特征向量的权重和所述相关性进行回归分析，通过所述数据预测模型中的输出层输出多个指标数据和对应的预测值。

5、根据权利要求 1 至 4 任意一项所述的方法，其特征在于，所述方法还包括：

根据预设频率从预设数据库中获取多个历史因素数据；

对多个历史因素数据进行聚类分析，得到分析结果；

根据所述分析结果进行特征选择，得到多个特征变量；

根据预设的算法计算多个特征变量的权重；及

根据多个特征变量和对应的权重对所述数据预测模型进行优化调整。

6、根据权利要求 1 所述的方法，其特征在于，所述方法还包括：

根据所述请求类型获取预设的集成函数；

根据所述预测结果数据中的多个预设时序参数和对应的预测值通过所述集成函数集成对应的视图预测数据；及

对所述视图预测数据添加事件类型标识和对应的接口调用参数，将所述视图预测数据发送至所述终端。

7、一种基于机器学习的数据预测处理装置，所述装置包括：

请求接收模块，用于接收终端发送的数据预测请求，所述数据预测请求包括请求类型和请求信息；

数据获取模块，用于识别所述请求信息中的字段因子，根据所述请求类型和所述字段因子获取多个目标因素数据，所述目标因素数据包括多个字段数据；

特征处理模块，用于对所述目标因素数据进行特征处理和特征提取，计算所述多个字段数据的特征维度值，根据所述特征维度值提取所述多个字段数据对应的特征向量；

数据预测模块，用于根据所述请求类型获取预设的数据预测模型，通过所述数据分析模型对所述特征变量进行训练和分析，得到多个指标数据和对应的预测值；及

结果数据生成模块，用于利用所述多个指标数据和对应的预测值生成的预测结果数据，并将所述预测结果数据推送至所述终端。

8、根据权利要求 7 所述的装置，其特征在于，所述特征处理模块还用于对所述目标因素数据对应的多个字段数据进行向量化，得到多个字段数据对应的特征向量；对所述特征变量进行衍生处理，得到处理后的多个特征向量；根据预设目标函数计算所述特征向量的多个特征维度值；根据预设距离算法和所述特征维度值计算多个特征向量之间的相似度；及提取出所述相似度达到预设阈值的特征向量。

9、根据权利要求 7 所述的装置，其特征在于，所述装置还包括模型训练模块，用于获取多个历史因素数据，根据所述多个历史因素数据生成对应的训练集数据和验证集数据；对所述训练集数据中的多个历史因素数据进行聚类分析，得到聚类结果；对所述聚类结果进行特征提取，提取出多个特征变量；获取预设的神经网络模型，通过所述神经网络模型对所述训练集数据进行训练，得到多个特征变量对应的特征维度值和权重，根据多个特征变量对应的特征维度值和权重构建初始数据预测模型；及利用所述验证集数据对所述数据预测模型进行进一步训练和验证，直到所述验证集数据中满足预设阈值的数量达到预设比值时，停止训练，得到训练完成的数据预测模型。

10、根据权利要求 7 所述的装置，其特征在于，所述数据预测模块还用于将所述多个特征向量作为所述数据预测模型的输入，根据所述数据预测模型中的预设函数计算多个特征向量对应的权重；根据所述权重计算多个特征向量与目标特征数据之间的相关性；及根据所述特征向量的权重和所述相关性进行回归分析，并通过所述数据预测模型中的输出层输出多个指标数据和对应的预测值。

11、根据权利要求 7 所述的装置，其特征在于，所述装置还包括模型优化模块，用于根据预设频率从预设数据库中获取多个历史因素数据；对多个历史因素数据进行聚类分析，得到分析结果；根据所述分析结果进行特征选择，得到多个特征变量；根据预设的算法计算多个特征变量的权重；及根据多个特征变量和对应的权重对所述数据预测模型进行优化调整。

12、根据权利要求 7 所述的装置，其特征在于，所述装置还包括视图集成模块，用于根据所述请求类型获取预设的集成函数；根据所述预测结果数据中的多个预设时序参数和对应的预测值通过所述集成函数集成对应的视图预测数据；及对所述视图预测数据添加事件类型标识和对应的接口调用参数，将所述视图预测数据发送至所述终端。

13、一种计算机设备，包括存储器及一个或多个处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器

执行以下步骤：

接收终端发送的数据预测请求，所述数据预测请求包括请求类型和请求信息；

识别所述请求信息中的字段因子，根据所述请求类型和所述字段因子获取多个目标因素数据，所述目标因素数据包括多个字段数据；

对所述目标因素数据进行特征处理和特征提取，计算所述多个字段数据的特征维度值，根据所述特征维度值提取所述多个字段数据对应的特征向量；

根据所述请求类型获取预设的数据预测模型，通过所述数据分析模型对所述特征变量进行训练和分析，得到多个指标数据和对应的预测值；及

利用所述多个指标数据和对应的预测值生成的预测结果数据，将所述预测结果数据推送至所述终端。

14、根据权利要求 13 所述的计算机设备，其特征在于，所述处理器执行计算机可读指令时还执行以下步骤：将所述多个特征向量作为所述数据预测模型的输入，根据所述数据预测模型中的预设函数计算多个特征向量对应的权重；根据所述权重计算多个特征向量与目标特征数据之间的相关性；及根据所述特征向量的权重和所述相关性进行回归分析，通过所述数据预测模型中的输出层输出多个指标数据和对应的预测值。

15、根据权利要求 13 所述的计算机设备，其特征在于，所述处理器执行计算机可读指令时还执行以下步骤：根据预设频率从预设数据库中获取多个历史因素数据；对多个历史因素数据进行聚类分析，得到分析结果；根据所述分析结果进行特征选择，得到多个特征变量；根据预设的算法计算多个特征变量的权重；及根据多个特征变量和对应的权重对所述数据预测模型进行优化调整。

16、根据权利要求 13 所述的计算机设备，其特征在于，所述处理器执行计算机可读指令时还执行以下步骤：根据所述请求类型获取预设的集成函数；根据所述预测结果数据中的多个预设时序参数和对应的预测值通过所述集成函数集成对应的视图预测数据；及对所述视图预测数据添加事件类型标识和对应的接口调用参数，将所述视图预测数据发送至所述终端。

17、一种非易失性的计算机可读存储介质，所述计算机可读存储介质中存储有至少一条计算机可读指令，所述计算机可读指令由处理器加载并执行以下步骤：

接收终端发送的数据预测请求，所述数据预测请求包括请求类型和请求信息；

识别所述请求信息中的字段因子，根据所述请求类型和所述字段因子获取多个目标因素数据，所述目标因素数据包括多个字段数据；

对所述目标因素数据进行特征处理和特征提取，计算所述多个字段数据的特征维度值，

根据所述特征维度值提取所述多个字段数据对应的特征向量；

根据所述请求类型获取预设的数据预测模型，通过所述数据分析模型对所述特征变量进行训练和分析，得到多个指标数据和对应的预测值；及

利用所述多个指标数据和对应的预测值生成的预测结果数据，将所述预测结果数据推送至所述终端。

18、根据权利要求 17 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还执行以下步骤：将所述多个特征向量作为所述数据预测模型的输入，根据所述数据预测模型中的预设函数计算多个特征向量对应的权重；根据所述权重计算多个特征向量与目标特征数据之间的相关性；及根据所述特征向量的权重和所述相关性进行回归分析，通过所述数据预测模型中的输出层输出多个指标数据和对应的预测值。

19、根据权利要求 17 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还执行以下步骤：根据预设频率从预设数据库中获取多个历史因素数据；对多个历史因素数据进行聚类分析，得到分析结果；根据所述分析结果进行特征选择，得到多个特征变量；根据预设的算法计算多个特征变量的权重；及根据多个特征变量和对应的权重对所述数据预测模型进行优化调整。

20、根据权利要求 17 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还执行以下步骤：根据所述请求类型获取预设的集成函数；根据所述预测结果数据中的多个预设时序参数和对应的预测值通过所述集成函数集成对应的视图预测数据；及对所述视图预测数据添加事件类型标识和对应的接口调用参数，将所述视图预测数据发送至所述终端。

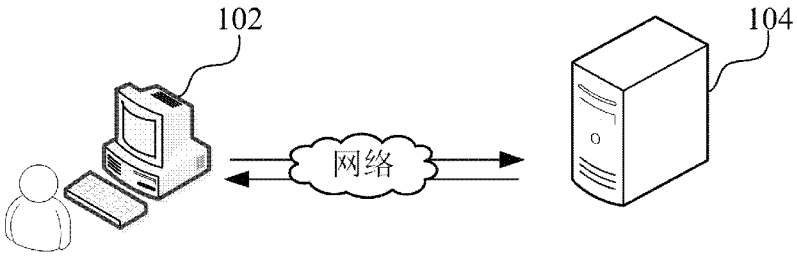
102
104
网络

图 1

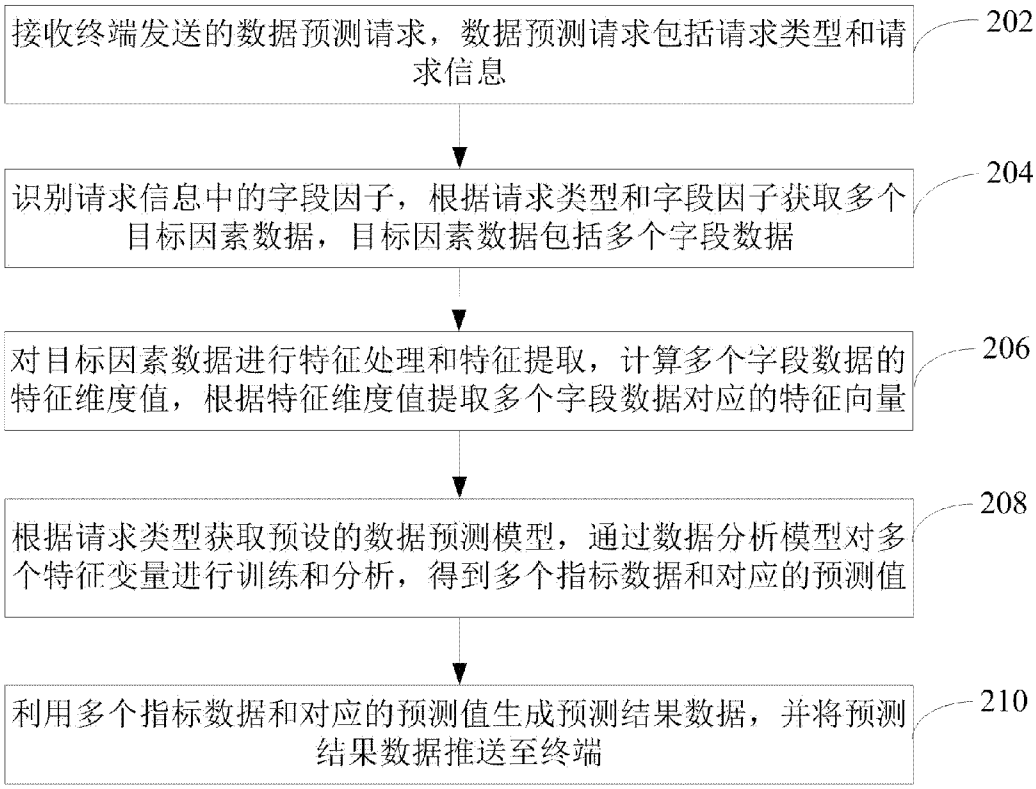
接收终端发送的数据预测请求，数据预测请求包括请求类型和请求信息
202
识别请求信息中的字段因子，根据请求类型和字段因子获取多个目标因素数据，目标因素数据包括多个字段数据
204
对目标因素数据进行特征处理和特征提取，计算多个字段数据的特征维度值，根据特征维度值提取多个字段数据对应的特征向量
206
根据请求类型获取预设的数据预测模型，通过数据分析模型对多个特征变量进行训练和分析，得到多个指标数据和对应的预测值
208
利用多个指标数据和对应的预测值生成预测结果数据，并将预测结果数据推送至终端
210

图 2

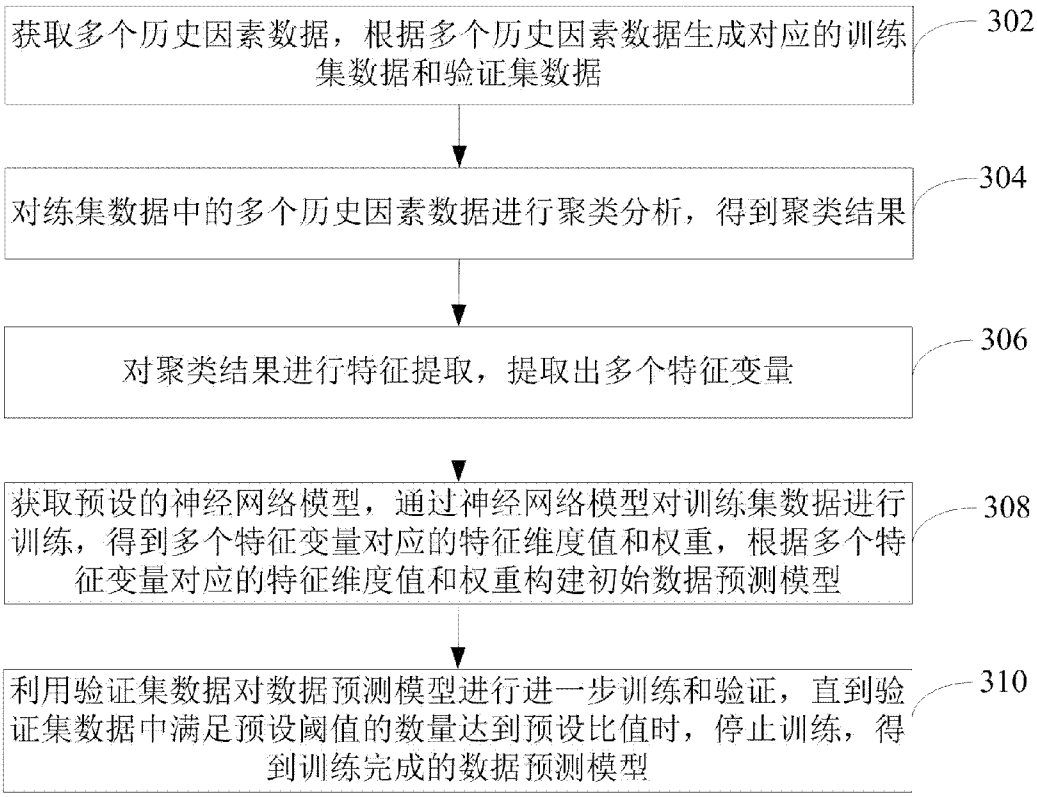
获取多个历史因素数据，根据多个历史因素数据生成对应的训练集数据和验证集数据
302
对练集数据中的多个历史因素数据进行聚类分析，得到聚类结果
304
对聚类结果进行特征提取，提取出多个特征变量
306
获取预设的神经网络模型，通过神经网络模型对训练集数据进行训练，得到多个特征变量对应的特征维度值和权重，根据多个特征变量对应的特征维度值和权重构建初始数据预测模型
308
利用验证集数据对数据预测模型进行进一步训练和验证，直到验证集数据中满足预设阈值的数量达到预设比值时，停止训练，得到训练完成的数据预测模型
310

图 3

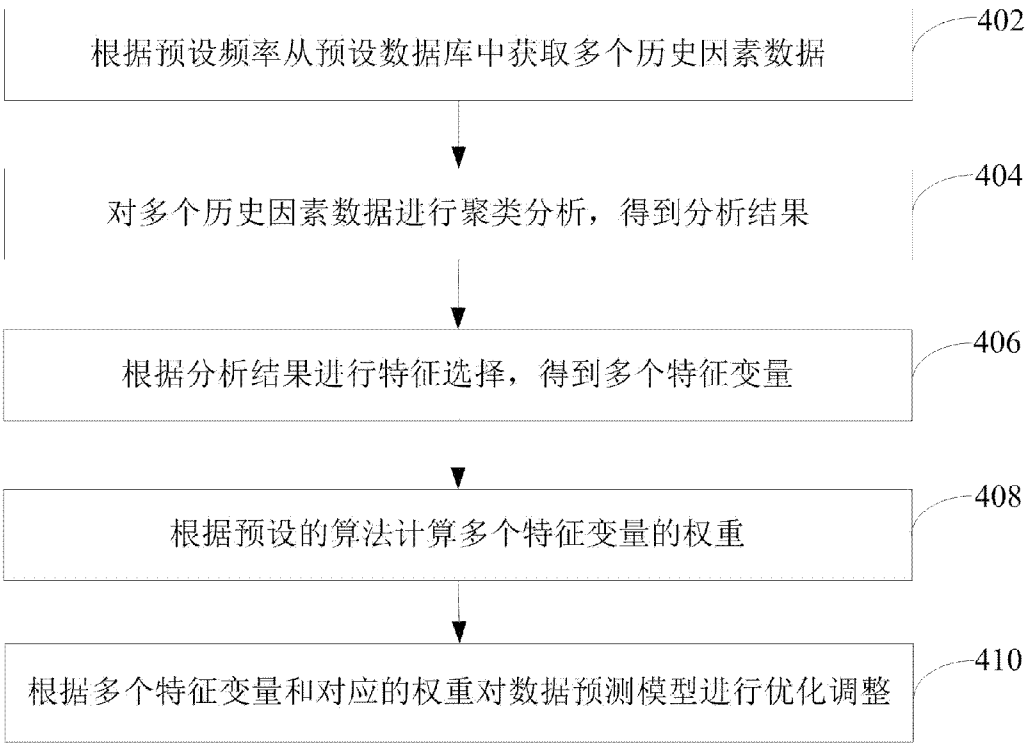
根据预设频率从预设数据库中获取多个历史因素数据
402
对多个历史因素数据进行聚类分析，得到分析结果
404
根据分析结果进行特征选择，得到多个特征变量
406
根据预设的算法计算多个特征变量的权重
408
根据多个特征变量和对应的权重对数据预测模型进行优化调整
410

图 4

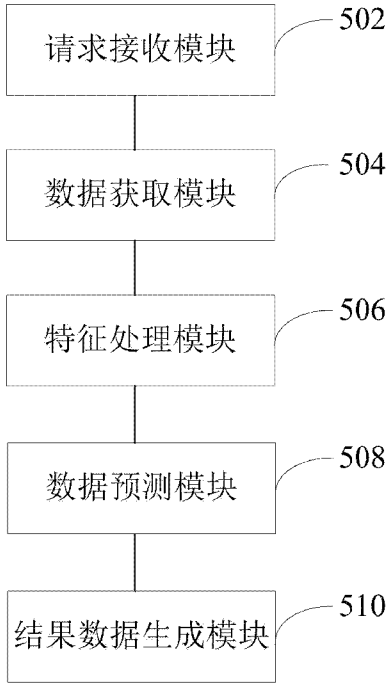
请求接收模块
502
数据获取模块
504
特征处理模块
506
数据预测模块
508
结果数据生成模块
510

图 5

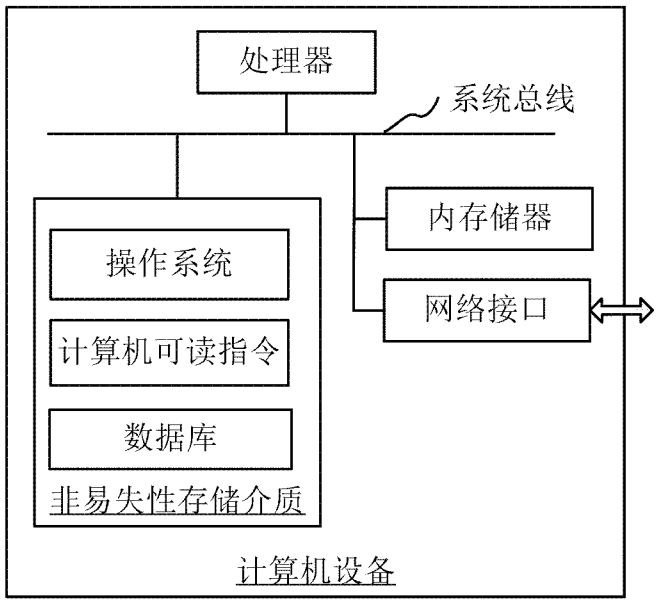
处理器
系统总线
内存储器
操作系统
网络接口
计算机可读指令
数据库
非易失性存储介质
计算机设备

图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/117290

A. CLASSIFICATION OF SUBJECT MATTER

G16H 50/80(2018.01)i; 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)

G16H G06N G06Q G06F

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: 预测, 模型, 请求, 类型, 疾病, 机器学习, 特征, 字段, 因子, predict+, machine learning, model, request, type, feature, field, factor

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109935338 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 25 June 2019 (2019-06-25) claims 1-10	1-20
X	CN 108170909 A (PING AN PROPERTY & CASUALTY INSURANCE COMPANY OF CHINA) 15 June 2018 (2018-06-15) description, paragraphs [0018]-[0060]	1-20
A	CN 108182515 A (CHINA PINGAN PROPERTY INSURANCE CO., LTD.) 19 June 2018 (2018-06-19) entire document	1-20
A	CN 108830380 A (OPEN AI LAB) 16 November 2018 (2018-11-16) entire document	1-20
A	US 2018285771 A1 (DRVISION TECHNOLOGIES LLC) 04 October 2018 (2018-10-04) entire document	1-20

☐ 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	Date of mailing of the international search report
10 February 2020	**17 February 2020**

Name and mailing address of the ISA/CN	Authorized officer
China National Intellectual Property Administration (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/117290

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 109935338 A	25 June 2019	None	
CN 108170909 A	15 June 2018	None	
CN 108182515 A	19 June 2018	None	
CN 108830380 A	16 November 2018	None	
US 2018285771 A1	04 October 2018	None	

国际检索报告

国际申请号
PCT/CN2019/117290

A. 主题的分类

G16H 50/80(2018.01)i; 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)

G16H G06N G06Q G06F

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,EPODOC:预测，模型，请求，类型，疾病，机器学习，特征，字段，因子，predict+, machine learning, model, request, type, feature, field, factor

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 109935338 A (平安科技深圳有限公司) 2019年 6月 25日 (2019-06-25) 权利要求1-10	1-20
X	CN 108170909 A (中国平安财产保险股份有限公司) 2018年 6月 15日 (2018-06-15) 说明书第[0018]-[0060]段	1-20
A	CN 108182515 A (中国平安财产保险股份有限公司) 2018年 6月 19日 (2018-06-19) 全文	1-20
A	CN 108830380 A (开放智能机器上海有限公司) 2018年 11月 16日 (2018-11-16) 全文	1-20
A	US 2018285771 A1 (DRVISION TECHNOLOGIES LLC) 2018年 10月 4日 (2018-10-04) 全文	1-20

☐ 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	国际检索报告邮寄日期
2020年 2月 10日	2020年 2月 17日

ISA/CN的名称和邮寄地址	受权官员
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	李韧
传真号 (86-10)62019451	电话号码 (86-10)53961624

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/117290

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 109935338 A	2019年 6月 25日	无	
CN 108170909 A	2018年 6月 15日	无	
CN 108182515 A	2018年 6月 19日	无	
CN 108830380 A	2018年 11月 16日	无	
US 2018285771 A1	2018年 10月 4日	无	