

(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)



(10) 国际公布号
WO 2020/177365 A1

(51) 国际专利分类号:
G06F 16/2458 (2019.01) **G06Q 40/08** (2012.01)

(21) 国际申请号: PCT/CN2019/116126

(22) 国际申请日: 2019 年 11 月 7 日 (07.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910171606.4 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)。

(54) Title: DATA MINING-BASED SOCIAL INSURANCE DATA PROCESSING METHOD AND APPARATUS, AND COMPUTER DEVICE

(54) 发明名称: 基于数据挖掘的社保数据处理方法、装置和计算机设备

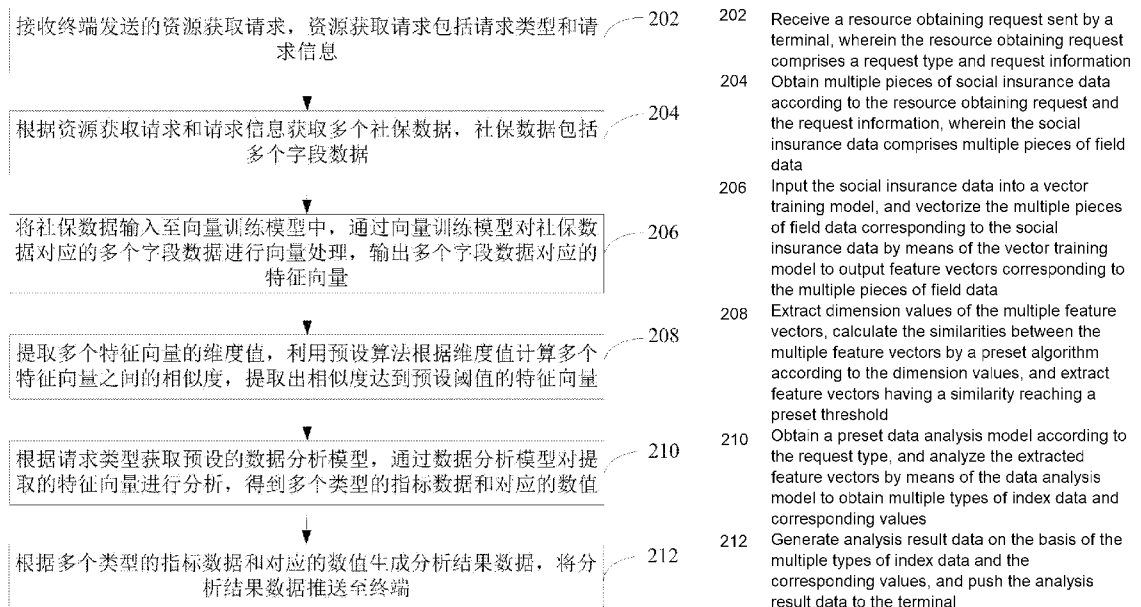


图 2

(57) Abstract: A data mining-based social insurance data processing method, comprising: receiving a resource obtaining request sent by a terminal, wherein the resource obtaining request comprises a request type and request information; obtaining multiple pieces of social insurance data according to the resource obtaining request and the request information, wherein the social insurance data comprises multiple pieces of field data; inputting the social insurance data into a vector training model, and vectorizing the multiple pieces of field data corresponding to the social insurance data by means of the vector training model to output feature vectors corresponding to the multiple pieces of field data; extracting dimension values of the multiple feature vectors, calculating the similarities between the

(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))。

multiple feature vectors by a preset algorithm according to the dimension values, and extracting feature vectors having a similarity reaching a preset threshold; obtaining a data analysis model according to the request type, and analyzing the extracted feature vectors by means of the data analysis model to obtain multiple types of index data and corresponding values; and generating analysis result data on the basis of the multiple types of index data and the corresponding values, and pushing the analysis result data to the terminal.

(57) 摘要: 一种基于数据挖掘的社保数据处理方法, 包括: 接收终端发送的资源获取请求, 资源获取请求包括请求类型和请求信息; 根据资源获取请求和请求信息获取多个社保数据, 社保数据包括多个字段数据; 将社保数据输入至向量训练模型, 通过向量训练模型对社保数据对应的多个字段数据进行向量处理, 输出多个字段数据对应的特征向量; 提取多个特征向量的维度值, 利用预设算法根据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; 根据请求类型获取数据分析模型, 通过数据分析模型对提取的特征向量进行分析, 得到多个类型的指标数据和对应的数值; 及根据多个类型的指标数据和对应的数值生成分析结果数据, 将分析结果数据推送至终端。

基于数据挖掘的社保数据处理方法、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2019101716064，申请名称为“基于数据挖掘的社保数据处理方法、装置和计算机设备”的中国专利申请案的优先权，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数据挖掘的社保数据处理方法、装置和计算机设备。

背景技术

随着经济的飞速发展，社会保险成为了民生经济的重要组成部分。随着计算机技术的不断发展，社会保险人员登记、社会保险金征收、社会保险金偿付等各个业务流程已经全部实现网络化和信息化，社保业务系统也积累了大量的社保数据。

然而，现有的对社保数据进行挖掘的方式中，大多的只是对社保数据进行查询以及简单的数据处理，对这些大量的社保数据没有进行更深层次的分析挖掘。且大量的社保数据具有数据量大，信息度繁杂冗余，在对大量的社保数据进行挖掘和分析时，大量的社保数据极易存在挖掘深度不够、流程混乱等情况，导致数据挖掘的效率和准确率较低。

发明内容

根据本申请公开的各种实施例一种基于数据挖掘的社保数据处理方法、装置和计算机设备。

一种基于数据挖掘的社保数据处理方法包括：

接收终端发送的资源获取请求，所述资源获取请求包括请求类型和请求信息；
根据所述资源获取请求和请求信息获取多个社保数据，所述社保数据包括多个字段数据；

将所述社保数据输入至向量训练模型中，通过向量训练模型对所述社保数据对应的多个字段数据进行向量处理，输出所述多个字段数据对应的特征向量；

提取所述多个特征向量的维度值，利用预设算法根据所述维度值计算多个特征

向量之间的相似度，提取出所述相似度达到预设阈值的特征向量；

根据所述请求类型获取预设的数据分析模型，通过所述数据分析模型对提取的特征向量进行分析，得到多个类型的指标数据和对应的数值；及

根据所述多个类型的指标数据和对应的数值生成分析结果数据，将所述分析结果数据推送至所述终端。

一种基于数据挖掘的社保数据处理装置包括：

请求接收模块，用于接收终端发送的资源获取请求，所述资源获取请求包括请求类型和请求信息；

数据获取模块，用于根据所述资源获取请求和请求信息获取多个社保数据，所述社保数据包括多个字段数据；

特征提取模块，用于将所述社保数据输入至向量训练模型中，通过向量训练模型对所述社保数据对应的多个字段数据进行向量处理，输出所述多个字段数据对应的特征向量；提取所述多个特征向量的维度值，利用预设算法根据所述维度值计算多个特征向量之间的相似度，提取出所述相似度达到预设阈值的特征向量；

数据分析模块，用于根据所述请求类型获取预设的数据分析模型，通过所述数据分析模型对提取的特征向量进行分析，得到多个类型的指标数据和对应的数值；及

数据推送模块，用于根据所述多个类型的指标数据和对应的数值生成分析结果数据，将所述分析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：

接收终端发送的资源获取请求，所述资源获取请求包括请求类型和请求信息；

根据所述资源获取请求和请求信息获取多个社保数据，所述社保数据包括多个字段数据；

将所述社保数据输入至向量训练模型中，通过向量训练模型对所述社保数据对应的多个字段数据进行向量处理，输出所述多个字段数据对应的特征向量；

提取所述多个特征向量的维度值，利用预设算法根据所述维度值计算多个特征向量之间的相似度，提取出所述相似度达到预设阈值的特征向量；

根据所述请求类型获取预设的数据分析模型，通过所述数据分析模型对提取的特征向量进行分析，得到多个类型的指标数据和对应的数值；及

根据所述多个类型的指标数据和对应的数值生成分析结果数据，将所述分析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：

接收终端发送的资源获取请求，所述资源获取请求包括请求类型和请求信息；

根据所述资源获取请求和请求信息获取多个社保数据，所述社保数据包括多个字段数据；

将所述社保数据输入至向量训练模型中，通过向量训练模型对所述社保数据对应的多个字段数据进行向量处理，输出所述多个字段数据对应的特征向量；

提取所述多个特征向量的维度值，利用预设算法根据所述维度值计算多个特征向量之间的相似度，提取出所述相似度达到预设阈值的特征向量；

根据所述请求类型获取预设的数据分析模型，通过所述数据分析模型对提取的特征向量进行分析，得到多个类型的指标数据和对应的数值；及

根据所述多个类型的指标数据和对应的数值生成分析结果数据，将所述分析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数据挖掘的社保数据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数据挖掘的社保数据处理方法的流程示意图。

图 3 为根据一个或多个实施例中对社保数据对应的多个字段数据进行向量化步骤的流程示意图。

图 4 为根据一个或多个实施例中通过数据分析模型对提取的特征向量进行分析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数据挖掘的社保数据处理装置的框图。

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

为了使本申请的目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。

本申请提供的基于数据挖掘的社保数据处理方法，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 可以向服务器发送资源获取请求，资源获取请求包括请求类型和请求信息。服务器 104 接收终端发送的资源获取请求后，根据资源获取请求和携带的请求信息获取多个社保数据，社保数据中包括了多个字段数据。服务器 104 进而对社保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。服务器 104 根据预设算法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。服务器进一步获取预设的数据分析模型，通过数据分析模型对提取的特征向量进行分析，得到对应的分析结果数据，并将分析结果数据推送至对应的终端 102。通过对大量的社保数据进行特征提取和筛查，并利用数据分析模型提取出的有价值的特征向量进行分析，由此能够有效地挖掘出社保数据中有价值的信息，进而有效地提高了社保数据的分析效率和准确率。

在其中一个实施例中，如图 2 所示，提供了一种基于数据挖掘的社保数据处理方法，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，接收终端发送的资源获取请求，资源获取请求包括请求类型和请求信息。

用户可以通过对应的终端输入相关的字段信息，并向服务器发送数据分析请求，资源获取请求可以是获取对社保数据进行分析后的结果数据。资源获取请求中携带了请求类型和请求信息，其中，请求类型可以是获取的资源数据的类型，例如社保类的分析数据。请求信息可以是用户输入的字段信息，例如可以是社保数据的范围、时间区间等字段信息。

步骤 204，根据资源获取请求和请求信息获取多个社保数据，社保数据包括多

个字段数据。

社保数据可以是社会保险数据，例如可以包括养老保险数据、医疗保险数据、失业保险数据、工伤保险数据以及生育保险数据等。服务器接收到终端发送的资源获取请求后，根据资源获取请求和请求信息从本地数据库或第三方数据库中获取多个社保数据。例如，当请求信息中获取的社保数据的范围为某个企业时，服务器则获取该企业对应的社保数据。社保数据中包括了多个字段数据，例如姓名、性别、年龄、地区、所属企业、缴费时长、缴费金额等字段信息。

步骤 206，将社保数据输入至向量训练模型中，通过向量训练模型对社保数据对应的多个字段数据进行向量处理，输出多个字段数据对应的特征向量。

服务器获取多个社保数据后，对社保数据对应的多个字段数据进行向量化。具体地，服务器可以获取预设的语料库，并根据社保数据从语料库中获取相关联的语料数据。服务器进一步获取预设的向量训练模型，例如，向量训练模型可以是基于 word2 vec 的神经网络模型。服务器则将社保数据和获取的相关联的语料数据输入至向量训练模型中，进而通过向量训练模型结合相关联的语料数据对社保数据进行词向量计算和训练，得到社保数据对应的多个词向量，并根据预设算法将词向量转换为对应的特征向量。由此能够得到多个字段数据对应的特征向量。

步骤 208，提取多个特征向量的维度值，利用预设算法根据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。

服务器得到多个字段数据对应的特征向量后，根据预设算法计算出多个特征向量之间的相似度。具体地，服务器可以首先根据预设的目标函数计算多个特征向量的多个维度值，维度值可以是每个特征向量对应不同维度的特征值。服务器进一步跟进预设的距离算法和特征向量的维度值计算多个特征向量之间的相似度，进而提取出相似度达到预设阈值的特征向量。

步骤 210，根据请求类型获取预设的数据分析模型，通过数据分析模型对提取的特征向量进行分析，得到多个类型的指标数据和对应的数值。

服务器提取出特征向量后，则进一步根据请求类型获取对应预设的数据分析模型，数据分析模型中可以包括多个不同类型的数据分析模块，例如参保缴费率、缴费基数分析、企业经营状况等多个类型的指标数据分析模块。通过数据分析模型对提取的特征向量进行分析。

具体地，服务器可以首先用过数据分析模型计算出多个特征向量的分布值和字段饱和度，其中分布值可以是特征向量对应的字段数据的值，字段饱和度可以是特

征向量和字段数据对应多个预设指标数据的值的饱和程度。服务器进一步通过数据分析模型对多个特征向量进行统计筛查，提取达到预设饱和值的特征向量。服务器则根据预设的语义分析算法对提取出的特征向量进行语义分析，得到每个特征向量的权重，即特征向量的重要程度值。服务器进而根据特征向量的分布值、字段饱和度以及权重对多个特征向量进行分析，得到特征向量对应多个类型的指标数据和数值。

步骤 212，根据多个类型的指标数据和对应的数值生成分析结果数据，将分析结果数据推送至终端。

服务器生成分析结果数据后，进而根据每个特征向量对应多个类型的指标数据和对应的数值生成分析结果数据。则将分析结果数据推送至对应的终端。进一步地，服务器还可以将分析结果数据生成预设格式的视图数据，并将生成的视图数据推送至对应的终端，由此可以使得用户能够清楚地了解分析结果数据。

例如，当获取到的社保数据为某个企业或某个区域的社保数据时，则对获取的多个社保数据进行挖掘分析，可以有效地分析出参保缴费率、缴费基数分析、企业经营状况等指标数据。通过对大量的社保数据进行特征提取和筛查，并利用数据分析模型提取出的有价值的特征向量进行分析，由此能够有效地挖掘和分析出社保数据中有价值的信息，由此有效地提高了社保数据的分析效率和准确率。

上述基于数据挖掘的社保数据处理方法中，服务器接收终端发送的资源获取请求后，根据资源获取请求和携带的请求信息获取多个社保数据，社保数据中包括了多个字段数据。服务器进而将社保数据输入至向量训练模型中，通过向量训练模型对社保数据对应的多个字段数据进行向量化，输出多个字段数据对应的特征向量。提取多个特征向量的维度值，利用预设算法根据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。服务器进一步获取预设的数据分析模型，通过数据分析模型对提取的特征向量进行分析，得到多个类型的指标数据和对应的数值，根据多个类型的指标数据和对应的数值生成分析结果数据，并将分析结果数据推送至对应的终端。通过对大量的社保数据进行特征提取和筛查，并利用数据分析模型提取出的有价值的特征向量进行分析，由此能够有效地挖掘出社保数据中有价值的信息，进而有效地提高了社保数据的分析效率和准确率。

在其中一个实施例中，如图 3 所示，通过向量训练模型对社保数据对应的多个字段数据进行向量处理的步骤，具体包括以下内容：

步骤 302，获取预设的语料库，根据社保数据从所述语料库中获取相关联的语

料数据。

终端可以向服务器发送资源获取请求，资源获取请求中携带了请求类型和请求信息。服务器接收终端发送的资源获取请求请求后，根据资源获取请求和请求信息从本地数据库或第三方数据库中获取对应的多个社保数据，社保数据中包括多个字段数据。

服务器获取多个社保数据后，进而获取预设的语料库。语料库可以是预先设置的包括多种与社保相关的词汇或语句的语料库。

步骤 304，获取预设的向量训练模型，通过向量训练模型对社保数据和语料数据进行词向量计算和训练，得到对应的多个词向量。

步骤 306，根据预设算法将多个词向量转换为对应的特征向量。

服务器进一步获取预设的向量训练模型，将社保数据和语料数据输入至向量训练模型中，例如，向量训练模型可以是基于 word2 vec 的神经网络模型。通过向量训练模型对社保数据和语料数据进行计算和训练，得到多个社保数据对应的词向量。例如，通过词向量的训练，每个字都可以训练得到 n 维空间中的一个向量，如当 n 取 2 维时，则“身”对应向量是 $[0.5365654, 0.726268]$ ，对应的“份”对应的词向量可能是 $[0.52222458, 0.7511456]$ ，这两个向量的 $\cos$ 值也就是余弦距离非常近，对应到语义空间中的距离非常近，则表示“身份”是成词的。如果 n 取 100，则每个字转化成 100 维度的向量。通过词向量模型对社保数据进行向量化，能够准确有效地提取出社保数据中的词向量。

服务器提取出社保数据中的词向量后，则进一步根据预设算法将词向量转换为对应的特征向量。例如，可以利用预设的向量表示法将词向量转换为对应的特征向量。由此能够有效地提取出社保数据对应的特征向量。

在其中一个实施例中，利用预设算法根据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步骤，包括：根据预设的目标函数计算多个特征向量的多个维度值；根据预设的距离算法和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。

终端可以向服务器发送资源获取请求，资源获取请求中携带了请求类型和请求信息。服务器接收终端发送的资源获取请求请求后，根据资源获取请求和请求信息从本地数据库或第三方数据库中获取对应的多个社保数据，社保数据中包括多个字段数据。

服务器对社保数据对应的多个字段数据进行向量化，由此得到多个字段数据对

应的特征向量。服务器则进一步根据预设算法计算多个特征向量之间的相关性。具体地，服务器可以根据预设的目标函数计算多个特征向量的多个维度值，并根据预设的距离算法和维度值计算多个特征向量之间的相似度，进而提取出相似度达到预设阈值的特征向量。例如，预设的距离算法可以为欧式距离算法。

例如，欧式距离函数的计算公式可以如下：

$$P(A, B) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (i=1, 2, 3, \dots)$$

目标函数的表达式可以为：

$$B_k = \arg \min(P(A_i, B_i))$$

通过目标函数使得 $P(A_i, B_i)$ 的值最小。对 Max、Min、Mean 三个维度的值进行抽取，Max 就是抽取向量同一个维度的最大值，例如 $0.5 > 0.2 > 0.1$ （第一维度）， $0.7 > 0.5 > 0.2$ （第二维度），则 Max 对应的就是 $[0.5, 0.7]$ ，同样的 Min 对应的就是 $[0.1, 0.2]$ ，Mean 均值对应的就是 $[0.8/3, 1.4/3]$ ，然后再将这三个向量横向连接起来，通过 $3*n$ 维度的向量来表示。例如当 n 取 2 时，身表示为 $[0.5, 0.2]$ ，份表示为 $[0.1, 0.7]$ ，证表示为 $[0.2, 0.5]$ 。因此通过 Max、Min、Mean 三个维度的抽取，Max 就是抽取向量同一个维度的最大值，如 $0.5 > 0.2 > 0.1$ （第一维度）， $0.7 > 0.5 > 0.2$ （第二维度），则 Max 对应的就是 $[0.5, 0.7]$ ，同样的 Min 对应的就是 $[0.1, 0.2]$ ，mean 均值对应的就是 $[0.8/3, 1.4/3]$ ，然后再将这三个向量横向连接起来，因此短文本“身份证”就可以用 $[0.5, 0.7, 0.1, 0.2, 0.8/3, 1.4/3]$ 这一个 6 个维度的向量来表示。同样的，如果短文本是“保险说明”，同样也可以用一个 6 个维度的向量来表示。因此无论是长度为多少的短文本，都可以通过 $3*n$ 维度的向量来表示。文本之间的相似度，则可以通过对多个文本的多个维度对应的向量的欧式距离进行计算，就可以得到文本的相似度结果。

服务器计算出多个特征向量之间的相似度后，则进一步提取出相似度达到预设阈值的特征向量。通过预设的目标函数和距离算法计算多个特征向量之间的相似度，进而提取出相似度达到预设阈值的特征向量，由此能够有效地提取对社保数据进行特征提取。

在其中一个实施例中，根据请求类型获取预设的数据分析模型之前，还包括：获取多个样本社保数据，样本社保数据包括多个字段数据；对样本社保数据进行向量化处理，得到多个字段数据对应的特征向量；对多个特征向量进行聚类，根据聚类结果计算特征向量之间的相关性以及每个特征向量的权重，提取出满足条件阈值的特征向量；利用提取出的特征向量和对应的权重按照预设算法构建数据分析模型。

服务器获取预设的数据分析模型之前，还需要对构建出数据分析模型。具体地，

服务器可以预先获取大量的样本社保数据，服务器首先对样本社保数据进向量处理，从而可以得到样本社保数据中多个字段数据对应的特征向量。服务器对样本社保数据进向量化后，对社保数据进行特征提取。具体地，服务器可以通过预设的聚类算法对多个特征向量进行聚类分析，计算出特征向量之间的相关性以及每个特征向量的权重，进而提取出满足预设条件阈值的特征向量。服务器进而根据提取出的特征向量和对应的权值按照预设算法构建数据分析模型。其中，数据分析模型中可以包括多个不同类型的数据分析模块，例如参保缴费率、缴费基数分析、企业经营状况等多个类型的指标数据分析模块。通过对大陆的社保数据进行分析和特征提取，并利用提取出的有价值的特征向量构建数据分析模型，由此能够有效地提高数据分析模型的准确率。

在其中一个实施例中，如图 4 所示，通过数据分析模型对提取的特征向量进行分析的步骤，具体包括以下内容：

步骤 402，通过数据分析模型计算出多个特征向量的分布值和字段饱和度。

步骤 404，对多个特征向量进行特征字段筛查，提取达到预设饱和值的特征向量。

终端可以向服务器发送资源获取请求，资源获取请求中携带了请求类型和请求信息。服务器接收终端发送的资源获取请求请求后，根据资源获取请求和请求信息从本地数据库或第三方数据库中获取对应的多个社保数据，社保数据中包括多个字段数据。

服务器对社保数据对应的多个字段数据进行向量化，由此得到多个字段数据对应的特征向量。服务器进而根据预设算法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。

服务器对社保数据进行特征提取，提取出对应的特征向量后，则进一步根据资源获取请求中的请求类型获取预设的数据分析模型，通过数据分析模型对提取的特征向量进行分析。具体地，服务器获取预设的数据分析模型后，将提取出的字段数据对应的特征向量输入至数据分析模型中，通过数据分析模型计算出字段数据的分布值和字段饱和度，并对字段数据进行特征字段筛查，例如可以是对字段数据进行统计筛查，以提取达到预设饱和值的特征向量。分布值可以是特征向量对应的字段数据的值。

例如，当某一个字段是年龄时，字段数据的分布值可以是 10-20、20-30、30-40 等每一个年龄段的人数分布。字段饱和度可以是特征向量和字段数据对应多个预设

指标数据的值的饱和程度，例如输入数据可能会存在一些不饱和的情况，如说有一些字段是空的，则字段数据的字段饱和度就比较低。因此，服务器需要对字段数据对应的特征向量进行统计探查进行二次字段筛选。

步骤 406，根据预设的语义分析算法，对提取出的特征向量进行分析，得到特征向量的权重。

步骤 408，根据特征向量的分布值和字段饱和度以及权重进行分析，得到特征向量对应多个类型的指标数据和对应的数值。

步骤 410，根据多个类型的指标数据和对应的数值生成分析结果数据。

服务器对多个特征向量进行统计筛查，提取达到预设饱和值的特征向量后，进一步根据预设的语义分析算法，对提取出的字段数据进行分析，得到字段数据对应的权重，即重要程度值。

服务器则根据字段数据的分布值和字段饱和度以及重要程度值进行分析，得到多个类型的指标数据和对应的数值，并根据多个类型的指标数据和对应的数值生成对应的分析结果数据。通过数据分析模型对提取的字段数据进行分析，由此有效地分析出社保数据对应的分析结果数据。

例如，语义分析可以是基于用户输入的字段与真实字段之间的匹配关系，请求信息中即包括用户输入的字段。如基于社保大数据上千维度的字段，包括脱敏身份证号、身高、体重、社保脱敏账号、社保属性等等，而用户可能只对自己特定的几个字段感兴趣。因此，用户只用输入感兴趣的字段，服务器则通过对提取出的社保数据对应的特征向量进行语义分析，分析出数据集中与用户输入的感兴趣字段相关的字段信息，并计算出特征向量对应的权重，进而获取相关联的字段信息。如果用户输入了一个较为模糊的感兴趣字段时，如“赔付”，而“赔付”中包含了年赔付次数、赔付金额、赔付原因等信息。

数据分析模型中可以包括多个不同类型的数据分析模块，例如参保缴费率、缴费基数分析、企业经营状况等多个类型的指标数据分析模块。服务器进而根据特征向量的分布值、字段饱和度以及权重对多个特征向量进行分析，得到特征向量对应多个类型的指标数据和数值。服务器进而根据每个特征向量对应多个类型的指标数据和对应的数值生成分析结果数据。服务器生成分析结果数据后，则将分析结果数据推送至对应的终端。通过对大量的社保数据进行特征提取和筛查，并利用数据分析模型提取出的有价值的特征向量进行分析，由此能够有效地挖掘和分析出社保数据中有价值的信息，由此有效地提高了社保数据的分析效率和准确率。

在其中一个实施例中，分析结果数据中包括多个类型的指标数据和对应的数值，该方法还包括：根据指标数据和对应的数值生成对应的指标分析数据；将指标分析数据按照预设方式生成对应的分析视图数据；对分析视图数据添加事件类型标识和对应的接口调用参数；将分析视图数据推送至终端。

服务器接收终端发送的资源获取请求后，根据资源获取请求和请求信息从本地数据库或第三方数据库中获取对应的多个社保数据，社保数据中包括多个字段数据。服务器则对社保数据对应的多个字段数据进行向量化，由此得到多个字段数据对应的特征向量。服务器进而根据预设算法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。

服务器对社保数据进行特征提取，提取出对应的特征向量后，则进一步根据资源获取请求中的请求类型获取预设的数据分析模型，通过数据分析模型对提取的特征向量进行分析。数据分析模型中可以包括多个不同类型的数据分析模块，例如参保缴费率、缴费基数分析、企业经营状况等多个类型的指标数据分析模块。服务器进而根据特征向量的分布值、字段饱和度以及权重对多个特征向量进行分析，得到特征向量对应多个类型的指标数据和数值。服务器进而根据每个特征向量对应多个类型的指标数据和对应的数值生成分析结果数据。

服务器通过对社保数据进行挖掘分析得到对应的分析结果数据后，分析结果数据中包括多个类型的指标数据和对应的数值。服务器还可以进一步根据指标数据类型将分析结果数据生成对应的多个指标类型的指标分析数据。服务器还可以将多个指标类型的模块数据按照预设方式分别生成对应可视化的分析视图数据。具体地，服务器可以根据请求类型获取预设的集成函数，根据分析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资源数据，并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。例如，预设的集成函数可以为 python 可视化函数，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。

服务器根据分析结果数据中的多个类型的指标数据和对应的数值通过集成函数集成对应的分析视图数据后，进一步对分析视图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分析视图数据进行调用，由此使得服务器或终端再次获取相关联的社保分析数据或分析视图数据时，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数据，进而提高了对社保数据的分析效率和利用价值。

服务器生成对应的分析视图数据后，则将分析视图数据发送至对应的终端，以使得对应的终端可以有效地根据挖掘出的社保数据结合对应的业务进行进一步分析，由此能够有效地对挖掘分析后的分析数据进行利用，从而有效提高了社保数据的挖掘效率和分析效率。

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，提供了一种基于数据挖掘的社保数据处理装置，包括：请求接收模块 502、数据获取模块 504、特征提取模块 506、数据分析模块 508 和数据推送模块 510，其中：

请求接收模块 502，用于接收终端发送的资源获取请求，资源获取请求包括请求类型和请求信息；

数据获取模块 504，用于根据资源获取请求和请求信息获取多个社保数据，社保数据包括多个字段数据；

特征提取模块 506，用于将社保数据输入至向量训练模型中，通过向量训练模型对社保数据对应的多个字段数据进行向量处理，输出多个字段数据对应的特征向量；提取多个特征向量的维度值，利用预设算法根据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；

数据分析模块 508，用于根据请求类型获取预设的数据分析模型，通过数据分析模型对提取的特征向量进行分析，得到多个类型的指标数据和对应的数值；

数据推送模块 510，用于根据多个类型的指标数据和对应的数值生成分析结果数据，将分析结果数据推送至终端。

在其中一个实施例中，特征提取模块 506 还用于获取预设的语料库，根据社保数据从语料库中获取相关联的语料数据；获取预设的向量训练模型，通过向量训练模型对社保数据和语料数据进行词向量计算和训练，得到对应的多个词向量；根据预设算法将词向量转换为对应的特征向量。

在其中一个实施例中，特征提取模块 506 还用于根据预设的目标函数计算多个

特征向量的多个维度值；根据预设的距离算法和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，该装置还包括模型建立模块，用于获取多个样本社保数据，样本社保数据包括多个字段数据；对样本社保数据进行向量化处理，得到多个字段数据对应的特征向量；对多个特征向量进行聚类，根据聚类结果计算特征向量之间的相关性以及每个特征向量的权重，提取出满足条件阈值的特征向量；及利用提取出的特征向量和对应的权重按照预设算法构建数据分析模型。

在其中一个实施例中，数据分析模块 508 还用于通过数据分析模型计算出多个特征向量的分布值和字段饱和度；对多个特征向量进行特征字段筛查，提取达到预设饱和值的特征向量；根据预设的语义分析算法，对提取出的特征向量进行语义分析，得到特征向量的权重；根据特征向量的分布值和字段饱和度以及权重进行分析，得到特征向量对应多个类型的指标数据和对应的数值；根据多个类型的指标数据和对应的数值生成分析结果数据。

在其中一个实施例中，分析结果数据中包括多个类型的指标数据和对应的数值，该装置还包括视图数据生成模块，用于根据指标数据和对应的数值生成对应的指标分析数据；将指标分析数据集成函数集成社保数据生成对应的分析视图数据；对分析视图数据添加事件类型标识和对应的接口调用参数，接口调用参数用于根据事件类型标识调用所生成的分析视图数据；将分析视图数据推送至终端。

关于基于数据挖掘的社保数据处理装置的具体限定可以参见上文中对于基于数据挖掘的社保数据处理方法的限定，在此不再赘述。上述基于数据挖掘的社保数据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社保数据、语料库和分析结果数据等数据。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

供的基于数据挖掘的社保数据处理方法的步骤。

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：

接收终端发送的资源获取请求，资源获取请求包括请求类型和请求信息；

根据资源获取请求和请求信息获取多个社保数据，社保数据包括多个字段数据；

将社保数据输入至向量训练模型中，通过向量训练模型对社保数据对应的多个字段数据进行向量处理，输出多个字段数据对应的特征向量；

提取多个特征向量的维度值，利用预设算法根据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；

根据请求类型获取预设的数据分析模型，通过数据分析模型对提取的特征向量进行分析，得到多个类型的指标数据和对应的数值；及

根据多个类型的指标数据和对应的数值生成分析结果数据，将分析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：

接收终端发送的资源获取请求，资源获取请求包括请求类型和请求信息；

根据资源获取请求和请求信息获取多个社保数据，社保数据包括多个字段数据；

将社保数据输入至向量训练模型中，通过向量训练模型对社保数据对应的多个字段数据进行向量处理，输出多个字段数据对应的特征向量；

提取多个特征向量的维度值，利用预设算法根据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；

根据请求类型获取预设的数据分析模型，通过数据分析模型对提取的特征向量进行分析，得到多个类型的指标数据和对应的数值；及

根据多个类型的指标数据和对应的数值生成分析结果数据，将分析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、一种基于数据挖掘的社保数据处理方法，所述方法包括：

接收终端发送的资源获取请求，所述资源获取请求包括请求类型和请求信息；
根据所述资源获取请求和请求信息获取多个社保数据，所述社保数据包括多个字段数据；

将所述社保数据输入至向量训练模型中，通过向量训练模型对所述社保数据对应的多个字段数据进行向量处理，输出所述多个字段数据对应的特征向量；

提取所述多个特征向量的维度值，利用预设算法根据所述维度值计算多个特征向量之间的相似度，提取出所述相似度达到预设阈值的特征向量；

根据所述请求类型获取预设的数据分析模型，通过所述数据分析模型对提取的特征向量进行分析，得到多个类型的指标数据和对应的数值；及

根据所述多个类型的指标数据和对应的数值生成分析结果数据，将所述分析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，所述通过向量训练模型对所述社保数据对应的多个字段数据进行向量处理的步骤，包括：

获取预设的语料库，根据所述社保数据从所述语料库中获取相关联的语料数据；

获取预设的向量训练模型，通过所述向量训练模型对所述社保数据和所述语料数据进行词向量计算和训练，得到对应的多个词向量；及

根据预设算法将所述多个词向量转换为对应的特征向量。

3、根据权利要求1所述的方法，其特征在于，所述利用预设算法根据所述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步骤，包括：

根据预设的目标函数计算多个特征向量的多个维度值；

根据预设的距离算法和所述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。

4、根据权利要求1所述的方法，其特征在于，所述根据所述请求类型获取预设的数据分析模型之前，还包括：

获取多个样本社保数据，所述样本社保数据包括多个字段数据；

对所述样本社保数据进行向量化处理，得到多个字段数据对应的特征向量；

对多个特征向量进行聚类，根据聚类结果计算特征向量之间的相关性以及每个

特征向量的权重，提取出满足条件阈值的特征向量；及

利用提取出的特征向量和对应的权重按照预设算法构建数据分析模型。

5、根据权利要求1所述的方法，其特征在于，通过所述数据分析模型对提取的特征向量进行分析的步骤，包括：

通过所述数据分析模型计算出多个特征向量的分布值和字段饱和度；

对多个特征向量进行特征字段筛查，提取达到预设饱和值的特征向量；

根据预设的语义分析算法，对提取出的特征向量进行分析，得到特征向量的权重；

根据所述特征向量的分布值和字段饱和度以及权重进行分析，得到所述特征向量对应多个类型的指标数据和对应的数值；及

根据所述多个类型的指标数据和对应的数值生成分析结果数据。

6、根据权利要求1至5任意一项所述的方法，其特征在于，所述方法还包括：

根据所述指标数据和对应的数值生成对应的指标分析数据；

将所述指标分析数据按照预设集成函数集成所述社保数据对应的分析视图数据；

对所述分析视图数据添加事件类型标识和对应的接口调用参数，所述接口调用参数用于根据事件类型标识调用所生成的分析视图数据；及

将所述分析视图数据推送至所述终端。

7、一种基于数据挖掘的社保数据处理装置，所述装置包括：

请求接收模块，用于接收终端发送的资源获取请求，所述资源获取请求包括请求类型和请求信息；

数据获取模块，用于根据所述资源获取请求和请求信息获取多个社保数据，所述社保数据包括多个字段数据；

特征提取模块，用于将所述社保数据输入至向量训练模型中，通过向量训练模型对所述社保数据对应的多个字段数据进行向量处理，输出所述多个字段数据对应的特征向量；提取所述多个特征向量的维度值，利用预设算法根据所述维度值计算多个特征向量之间的相似度，提取出所述相似度达到预设阈值的特征向量；

数据分析模块，用于根据所述请求类型获取预设的数据分析模型，通过所述数据分析模型对提取的特征向量进行分析，得到多个类型的指标数据和对应的数值；及

数据推送模块，用于根据所述多个类型的指标数据和对应的数值生成分析结果

数据，将所述分析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，所述特征提取模块还用于获取预设的语料库，根据所述社保数据从所述语料库中获取相关联的语料数据；获取预设的向量训练模型，通过所述向量训练模型对所述社保数据和所述语料数据进行词向量计算和训练，得到对应的多个词向量；及根据预设算法将所述多个词向量转换为对应的特征向量。

9、根据权利要求 7 所述的装置，其特征在于，所述特征提取模块还用于根据预设的目标函数计算多个特征向量的多个维度值；根据预设的距离算法和所述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。

10、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建立模块，用于获取多个样本社保数据，所述样本社保数据包括多个字段数据；对所述样本社保数据进行向量化处理，得到多个字段数据对应的特征向量；对多个特征向量进行聚类，根据聚类结果计算特征向量之间的相关性以及每个特征向量的权重，提取出满足条件阈值的特征向量；及利用提取出的特征向量和对应的权重按照预设算法构建数据分析模型。

11、根据权利要求 7 所述的装置，其特征在于，所述数据分析模块还用于通过所述数据分析模型计算出多个特征向量的分布值和字段饱和度；对多个特征向量进行特征字段筛查，提取达到预设饱和值的特征向量；根据预设的语义分析算法，对提取出的特征向量进行分析，得到特征向量的权重；根据所述特征向量的分布值和字段饱和度以及权重进行分析，得到所述特征向量对应多个类型的指标数据和对应的数值；及根据所述多个类型的指标数据和对应的数值生成分析结果数据。

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数据生成模块，用于根据所述指标数据和对应的数值生成对应的指标分析数据；将所述指标分析数据按照预设集成函数集成所述社保数据对应的分析视图数据；对所述分析视图数据添加事件类型标识和对应的接口调用参数，所述接口调用参数用于根据事件类型标识调用所生成的分析视图数据；及将所述分析视图数据推送至所述终端。

13、一种计算机设备，包括存储器及一个或多个处理器，所述存储器存储有至少一条计算机可读指令，所述计算机可读指令由所述处理器加载并执行以下步骤：
接收终端发送的资源获取请求，所述资源获取请求包括请求类型和请求信息；
根据所述资源获取请求和请求信息获取多个社保数据，所述社保数据包括多个

字段数据；

将所述社保数据输入至向量训练模型中，通过向量训练模型对所述社保数据对应的多个字段数据进行向量处理，输出所述多个字段数据对应的特征向量；

提取所述多个特征向量的维度值，利用预设算法根据所述维度值计算多个特征向量之间的相似度，提取出所述相似度达到预设阈值的特征向量；

根据所述请求类型获取预设的数据分析模型，通过所述数据分析模型对提取的特征向量进行分析，得到多个类型的指标数据和对应的数值；及

根据所述多个类型的指标数据和对应的数值生成分析结果数据，将所述分析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：获取预设的语料库，根据所述社保数据从所述语料库中获取相关联的语料数据；获取预设的向量训练模型，通过所述向量训练模型对所述社保数据和所述语料数据进行词向量计算和训练，得到对应的多个词向量；及根据预设算法将所述多个词向量转换为对应的特征向量。

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：通过所述数据分析模型计算出多个特征向量的分布值和字段饱和度；对多个特征向量进行特征字段筛查，提取达到预设饱和值的特征向量；根据预设的语义分析算法，对提取出的特征向量进行分析，得到特征向量的权重；根据所述特征向量的分布值和字段饱和度以及权重进行分析，得到所述特征向量对应多个类型的指标数据和对应的数值；及根据所述多个类型的指标数据和对应的数值生成分析结果数据。

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：根据所述指标数据和对应的数值生成对应的指标分析数据；将所述指标分析数据按照预设集成函数集成所述社保数据对应的分析视图数据；对所述分析视图数据添加事件类型标识和对应的接口调用参数，所述接口调用参数用于根据事件类型标识调用所生成的分析视图数据；及将所述分析视图数据推送至所述终端。

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：

接收终端发送的资源获取请求，所述资源获取请求包括请求类型和请求信息；根据所述资源获取请求和请求信息获取多个社保数据，所述社保数据包括多个

字段数据；

将所述社保数据输入至向量训练模型中，通过向量训练模型对所述社保数据对应的多个字段数据进行向量处理，输出所述多个字段数据对应的特征向量；

提取所述多个特征向量的维度值，利用预设算法根据所述维度值计算多个特征向量之间的相似度，提取出所述相似度达到预设阈值的特征向量；

根据所述请求类型获取预设的数据分析模型，通过所述数据分析模型对提取的特征向量进行分析，得到多个类型的指标数据和对应的数值；及

根据所述多个类型的指标数据和对应的数值生成分析结果数据，将所述分析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：获取预设的语料库，根据所述社保数据从所述语料库中获取相关联的语料数据；获取预设的向量训练模型，通过所述向量训练模型对所述社保数据和所述语料数据进行词向量计算和训练，得到对应的多个词向量；及根据预设算法将所述多个词向量转换为对应的特征向量。

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：通过所述数据分析模型计算出多个特征向量的分布值和字段饱和度；对多个特征向量进行特征字段筛查，提取达到预设饱和值的特征向量；根据预设的语义分析算法，对提取出的特征向量进行分析，得到特征向量的权重；根据所述特征向量的分布值和字段饱和度以及权重进行分析，得到所述特征向量对应多个类型的指标数据和对应的数值；及根据所述多个类型的指标数据和对应的数值生成分析结果数据。

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：根据所述指标数据和对应的数值生成对应的指标分析数据；将所述指标分析数据按照预设集成函数集成所述社保数据对应的分析视图数据；对所述分析视图数据添加事件类型标识和对应的接口调用参数，所述接口调用参数用于根据事件类型标识调用所生成的分析视图数据；及将所述分析视图数据推送至所述终端。

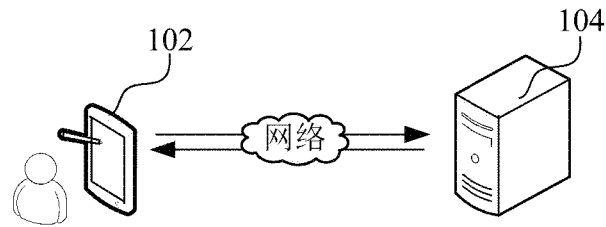


图 1

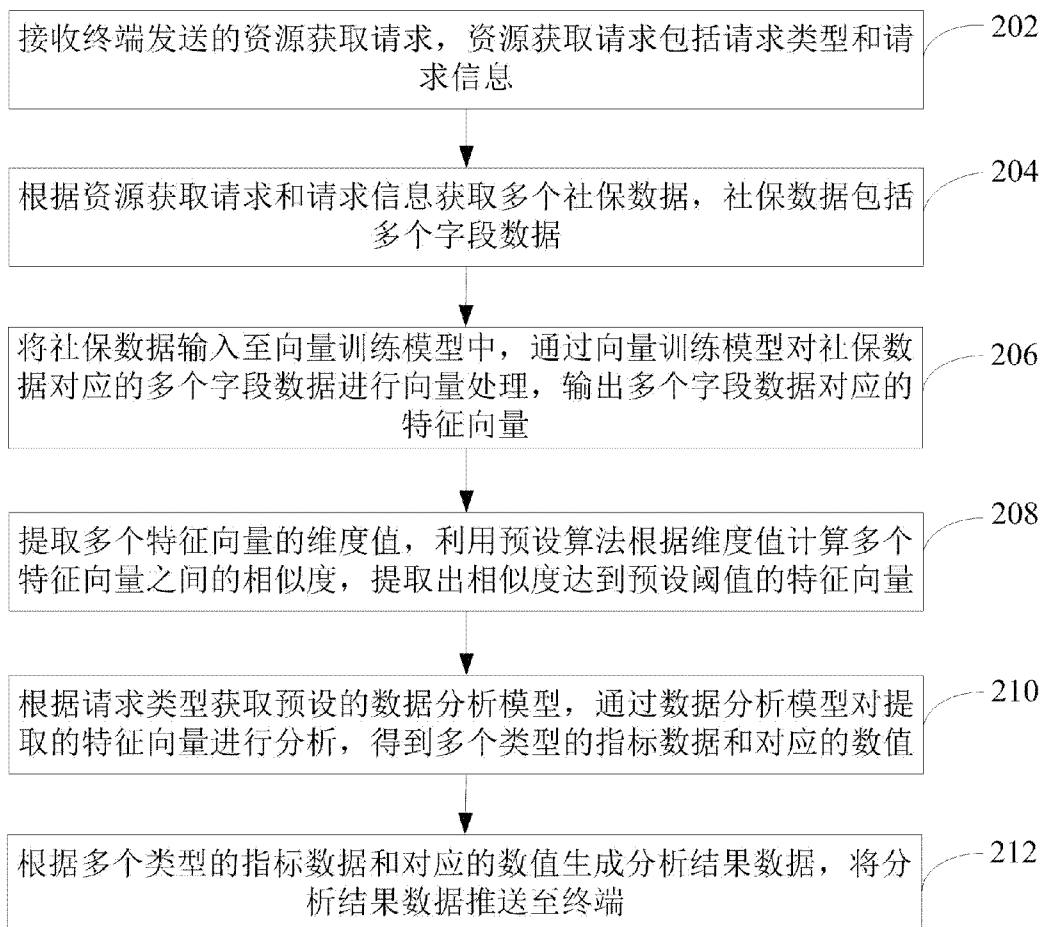


图 2

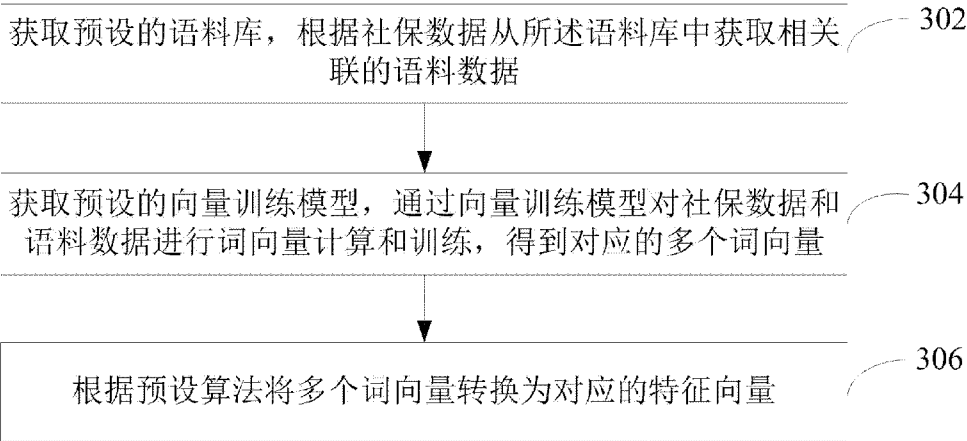


图 3



图 4

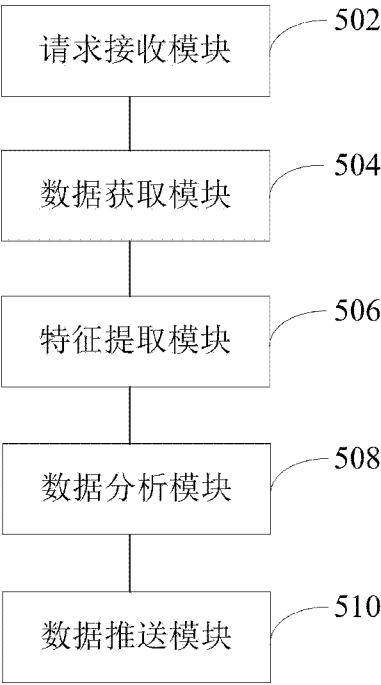


图 5

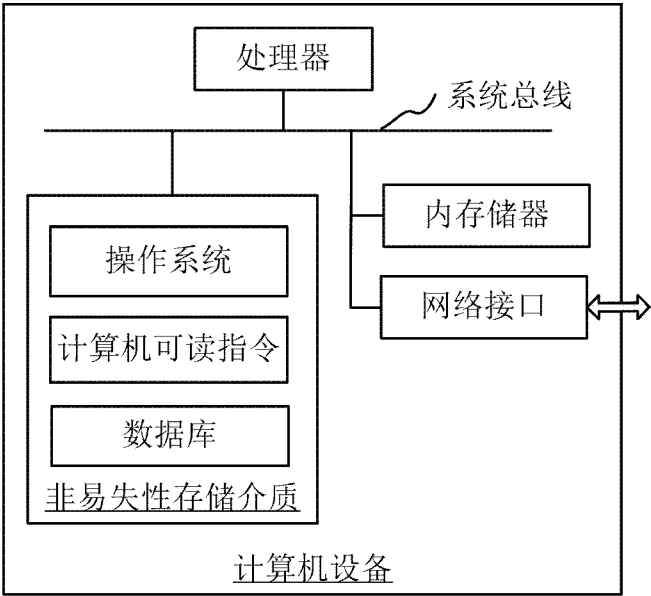


图 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/116126

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/2458(2019.01)i; G06Q 40/08(2012.01)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)

G06F; G06Q

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, GOOGLE: 保险, 社保, 向量, 矢量, 分析模型, 预测模型, 维度, 相似度, 相似性, 相关性, 相关度, 字段, 请求, request, vector, analysis, model, field, similarity, insurance

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110008250 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 12 July 2019 (2019-07-12) claims 1-10, and description, paragraph [0068]	1-20
X	CN 108520324 A (BEIJING JD FINANCE TECHNOLOGY CO., LTD.) 11 September 2018 (2018-09-11) claims 1-14, and description, paragraphs [0029]-[0105]	1-20
A	CN 105786711 A (GUANGZHOU HUADUO NETWORK TECHNOLOGY CO., LTD.) 20 July 2016 (2016-07-20) entire document	1-20
A	US 2012321202 A1 (FERTIK, M. B. S.) 20 December 2012 (2012-12-20) 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

27 December 2019

Date of mailing of the international search report

23 January 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/CN2019/116126

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110008250	A	12 July 2019	None			
CN	108520324	A	11 September 2018	None			
CN	105786711	A	20 July 2016	None			
US	2012321202	A1	20 December 2012	WO	2012177794	A2	27 December 2012
				US	2012321204	A1	20 December 2012
				US	2012321188	A1	20 December 2012

国际检索报告

国际申请号

PCT/CN2019/116126

A. 主题的分类 G06F 16/2458(2019.01)i; G06Q 40/08(2012.01)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) G06F; G06Q 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, GOOGLE: 保险, 社保, 向量, 矢量, 分析模型, 预测模型, 维度, 相似度, 相似性, 相关性, 相关度, 字段, 请求, request, vector, analysis, model, field, similarity, insurance																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 110008250 A (平安科技深圳有限公司) 2019年 7月 12日 (2019 - 07 - 12) 权利要求1-10, 说明书第[0068]段</td> <td>1-20</td> </tr> <tr> <td>X</td> <td>CN 108520324 A (北京京东金融科技控股有限公司) 2018年 9月 11日 (2018 - 09 - 11) 权利要求1-14, 说明书第[0029]-[0105]段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 105786711 A (广州华多网络科技有限公司) 2016年 7月 20日 (2016 - 07 - 20) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 2012321202 A1 (FERTIK, Michael Benjamin Selkove等) 2012年 12月 20日 (2012 - 12 - 20) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 110008250 A (平安科技深圳有限公司) 2019年 7月 12日 (2019 - 07 - 12) 权利要求1-10, 说明书第[0068]段	1-20	X	CN 108520324 A (北京京东金融科技控股有限公司) 2018年 9月 11日 (2018 - 09 - 11) 权利要求1-14, 说明书第[0029]-[0105]段	1-20	A	CN 105786711 A (广州华多网络科技有限公司) 2016年 7月 20日 (2016 - 07 - 20) 全文	1-20	A	US 2012321202 A1 (FERTIK, Michael Benjamin Selkove等) 2012年 12月 20日 (2012 - 12 - 20) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
PX	CN 110008250 A (平安科技深圳有限公司) 2019年 7月 12日 (2019 - 07 - 12) 权利要求1-10, 说明书第[0068]段	1-20															
X	CN 108520324 A (北京京东金融科技控股有限公司) 2018年 9月 11日 (2018 - 09 - 11) 权利要求1-14, 说明书第[0029]-[0105]段	1-20															
A	CN 105786711 A (广州华多网络科技有限公司) 2016年 7月 20日 (2016 - 07 - 20) 全文	1-20															
A	US 2012321202 A1 (FERTIK, Michael Benjamin Selkove等) 2012年 12月 20日 (2012 - 12 - 20) 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 2019年 12月 27日		国际检索报告邮寄日期 2020年 1月 23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 王怡轩 电话号码 86-10-53961621															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/116126

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	110008250	A	2019年 7月 12日	无			
CN	108520324	A	2018年 9月 11日	无			
CN	105786711	A	2016年 7月 20日	无			
US	2012321202	A1	2012年 12月 20日	WO	2012177794	A2	2012年 12月 27日
				US	2012321204	A1	2012年 12月 20日
				US	2012321188	A1	2012年 12月 20日