

(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 7 月 2 日 (02.07.2020)



(10) 国际公布号
WO 2020/132918 A1

(51) 国际专利分类号:
G16H 20/10 (2018.01)

(21) 国际申请号: PCT/CN2018/123761

(22) 国际申请日: 2018 年 12 月 26 日 (26.12.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811585289.2 2018 年 12 月 24 日 (24.12.2018) CN

(71) 申请人: 深圳市优必选科技有限公司 (UBTECH ROBOTICS CORP) [CN/CN]; 中国广东省深圳市南山区学苑大道 1001 号南山智园 C1 栋 16、22 楼, Guangdong 518000 (CN)。

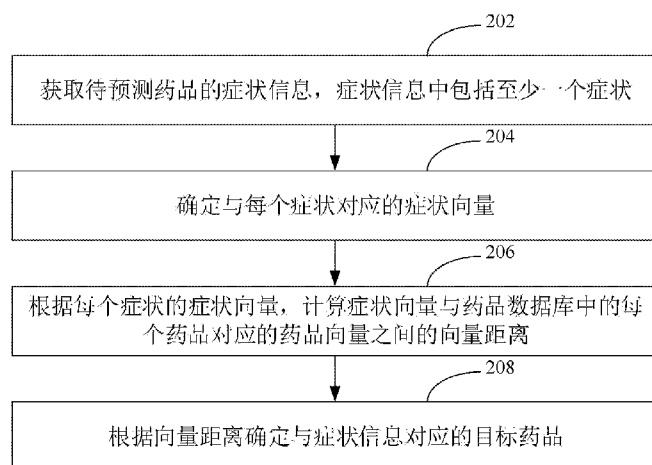
(72) 发明人: 熊友军 (XIONG, Youjun); 中国广东省深圳市南山区学苑大道 1001 号南山智园 C1 栋 16、22 楼, Guangdong 518000 (CN)。 罗沛鹏 (LUO, Peipeng); 中国广东省深圳市南山区学苑大道 1001 号南山智园 C1 栋 16、22 楼, Guangdong 518000 (CN)。 廖洪涛 (LIAO, Hongtao); 中国广东省深圳市南山区学苑大道 1001 号南山智园 C1 栋 16、22 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳中细软知识产权代理有限公司 (SHENZHEN CIPRUN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国广东省深圳市南山区粤海街道粤兴四道 1 号中山大学深圳产学研大楼 11 楼, Guangdong 518000 (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,

(54) **Title:** METHOD AND DEVICE FOR PHARMACEUTICAL FORECASTING, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 药品预测方法、装置、计算机设备及存储介质



202 ACQUIRE SYMPTOM INFORMATION OF MEDICAMENT TO BE FORECASTED, SYMPTOM INFORMATION COMPRISING AT LEAST ONE SYMPTOM
204 DETERMINE SYMPTOM VECTOR CORRESPONDING TO EACH SYMPTOM
206 CALCULATE, ON THE BASIS OF SYMPTOM VECTOR OF EACH SYMPTOM, VECTOR DISTANCE BETWEEN SYMPTOM VECTOR AND MEDICAMENT VECTORS CORRESPONDING TO MEDICAMENTS IN MEDICAMENT DATABASE
208 DETERMINE, ON THE BASIS OF VECTOR DISTANCE, TARGET MEDICAMENT CORRESPONDING TO SYMPTOM INFORMATION

图 2

(57) **Abstract:** A method and device for pharmaceutical forecasting, a computer device, and a storage medium. The method comprises: acquiring symptom information of medicament to be forecasted, the symptom information comprising at least one symptom (202); determining a symptom vector corresponding to each symptom (204); calculating, on the basis of the symptom vector of each symptom, the vector distance between the symptom vector and medicament vectors corresponding to medicaments in a medicament database (206); and determining, on the basis of the vector distance, a target medicament corresponding to the symptom information (208). The method for pharmaceutical forecasting greatly increases the efficiency of forecasting.

(57) **摘要:** 一种药品预测方法、装置、计算机设备及存储介质, 该方法包括: 获取待预测药品的症状信息, 所述症状信息中包括至少一个症状 (202); 确定与每个症状对应的症状向量 (204); 根据所述每个症状的症状向量, 计算所述症状向量与药品数据库中的每个药品对应的药品向量之间的向量距离 (206); 根据所述向量距离确定与所述症状信息对应的目标药品 (208), 该药品预测方法大大提高了预测的效率。

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))。

药品预测方法、装置、计算机设备及存储介质

技术领域

本发明涉及计算机处理领域，尤其是涉及一种药品预测方法、装置、计算机设备及存储介质。

背景技术

随着人工智能的兴起，智能问诊已经成为了一个趋势，根据症状对药品进行预测可以作为辅助医疗的技术，在一定程度上能减轻医院医生的压力。

但是，传统的机器问诊是通过文本匹配的方式直接根据症状去数据库里面搜药品的说明书，然后找到药品进行预测，这种文本匹配的方式十分僵化，预测药品的效率很低。

发明内容

基于此，有必要针对上述问题，提供了一种预测效率高的药品预测方法、装置、计算机设备及存储介质。

第一方面，本发明实施例提供一种药品预测方法，所述方法包括：
获取待预测药品的症状信息，所述症状信息中包括至少一个症状；
确定与每个症状对应的症状向量；

根据所述每个症状的症状向量，计算所述症状向量与药品数据库中的每个药品对应的药品向量之间的向量距离；

根据所述向量距离确定与所述症状信息对应的目标药品。

第二方面，本发明实施例提供一种药品预测装置，所述装置包括：

获取模块，用于获取待预测药品的症状信息，所述症状信息中包括至少一个症状；

向量确定模块，用于确定与每个症状对应的症状向量；

计算模块，用于根据所述每个症状的症状向量，计算所述症状向量与药品

数据库中的每个药品对应的药品向量之间的向量距离；

药品确定模块，用于根据所述向量距离确定与所述症状信息对应的目标药品。

第三方面，本发明实施例提供一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行如下步骤：

获取待预测药品的症状信息，所述症状信息中包括至少一个症状；

确定与每个症状对应的症状向量；

根据所述每个症状的症状向量，计算所述症状向量与药品数据库中的每个药品对应的药品向量之间的向量距离；

根据所述向量距离确定与所述症状信息对应的目标药品。

第四方面，本发明实施例提供一种计算机可读存储介质，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行如下步骤：

获取待预测药品的症状信息，所述症状信息中包括至少一个症状；

确定与每个症状对应的症状向量；

根据所述每个症状的症状向量，计算所述症状向量与药品数据库中的每个药品对应的药品向量之间的向量距离；

根据所述向量距离确定与所述症状信息对应的目标药品。

上述药品预测方法、装置、计算机设备及存储介质，在获取到待预测药品的症状信息后，确定与每个症状对应的症状向量，然后根据每个症状的症状向量，计算症状向量与药品数据库中每个药品的药品向量之间的向量距离，然后根据向量距离确定症状信息对应的目标药品。上述药品预测方法，通过将症状和药品之间的匹配关系转换为了向量运算，通过症状向量与药品向量之间的距离运算便可以快速找到目标药品，大大提高了查找的速度，提高了预测药品的效率，且该查找方式有利于提高预测的准确度。

附图说明

为了更清楚地说明本发明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发明的一些实施例，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他的附图。

5 其中：

图 1 为一个实施例中药品预测方法的应用环境图；

图 2 为一个实施例中药品预测方法的流程图；

图 3 为一个实施例中 CBOW 和 Skip-gram 预测的原理示意图；

图 4 为另一个实施例中药品预测方法的流程图；

10 图 5 为一个实施例中不同症状在二维空间的可视化示意图；

图 6 为一个实施例中词向量模型训练的方法流程示意图；

图 7 为一个实施例中确定目标药品的方法流程图；

图 8 为一个实施例中药品预测的方法流程示意图；

图 9 为一个实施例中药品预测装置的结构框图；

15 图 10 为另一个实施例中药品预测装置的结构框图；

图 11 为又一个实施例中药品预测装置的结构框图；

图 12 为一个实施例中计算机设备的内部结构图。

具体实施方式

20 为了使本发明的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本发明进行进一步详细说明。应当理解，此处所描述的具体实施例仅仅用以解释本发明，并不用于限定本发明。

图 1 为一个实施例中药品预测方法的应用环境图。参照图 1，该药品预测应用于药品预测系统。该药品预测系统包括终端 110 和服务器 120。终端 110 和服务器 120 通过网络连接，终端 110 具体可以是台式终端或移动终端，移动终端具体可以是手机、平板电脑、笔记本电脑等中的至少一种。服务器 120 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 110 用于

25

获取待预测药品的症状信息，症状信息中包括至少一个症状，然后将症状信息上传到服务器 120，服务器 120 获取到待预测药品的症状信息后，确定与每个症状对应的症状向量，根据所述每个症状的症状向量，计算所述症状向量与药品数据库中的每个药品对应的药品向量之间的向量距离，根据所述向量距离确定与所述症状信息对应的目标药品，将目标药品返回给终端 110。

在另一个实施例中，上述药品预测方法可以直接应用于终端 110，终端 110 用于获取待预测药品的症状信息，所述症状信息中包括至少一个症状，确定与每个症状对应的症状向量，根据所述每个症状的症状向量，计算所述症状向量与药品数据库中的每个药品对应的药品向量之间的向量距离，根据所述向量距离确定与所述症状信息对应的目标药品。

如图 2 所示，提出了一种药品预测方法，该药品预测方法可以应用于终端，也可以应用于服务器，本实施例中以应用于终端为例说明，该药品预测方法具体包括以下步骤：

步骤 202，获取待预测药品的症状信息，症状信息中包括至少一个症状。

其中，症状信息用于描述生病特征的信息，症状信息中包括一个或多个症状，症状是指生病的特征，比如，头痛、发烧等症状。为了给生病的人、动物以及植物预测出合适的药品，需要获取相应的人或动物或植物的症状信息，以便根据该症状信息进行药品的预测。

步骤 204，确定与每个症状对应的症状向量。

其中，症状向量是指症状的向量表示。症状向量可以通过词向量模型（比如，word2vec 模型）进行训练得到。在一个实施例中，通过将药品说明书的治疗症状进行分词，然后通过 word2vec 模型训练，word2vec 训练的词向量能够将词共现多的词聚类，所以当症状与症状在一起出现的多时，他们得到的词向量在空间是很接近的，比如，对于药品，一类症状往往会常常一起出现，比如，“头痛”、“发热”经常会一起出现。在通过词向量模型训练得到各个症状对应的症状向量后，将症状向量与症状进行关联存储，当获取到症状后，根据症状与症状向量的对应关系便可快速查找到对应的症状向量。

步骤 206, 根据每个症状的症状向量, 计算症状向量与药品数据库中的每个药品对应的药品向量之间的向量距离。

其中, 药品数据库中存储了每个药品以及每个药品对应的药品向量。通过计算症状向量与药品向量之间的空间向量距离来查找相应的药品。向量距离是指向量之间的距离, 在一个实施例中, 可以采用欧式距离来进行向量距离的运算。令 d 表示向量距离, X_{1i} 和 X_{2i} 分别表示症状向量和药品向量, 则相应的向量距离的计算公式如下: $d = \sqrt{\sum_{i=1}^N (X_{1i} - X_{2i})^2}$ 。计算得到的向量距离越近, 说明离药品的距离越近。

步骤 208, 根据向量距离确定与症状信息对应的目标药品。

其中, 在计算得到症状向量与每个药品的药品向量之间的向量距离后, 根据计算得到的向量距离从小到大进行排序, 将排序最靠前的药品作为预测的目标药品。当存在多个症状时, 分别计算每个症状向量与药品向量之间的向量距离, 然后将计算得到的多个向量距离进行平均确定与药品向量之间的平均向量距离。通过比较每个药品对应的平均向量距离来确定目标药品。

在一个实施例中, 如果是多个症状预测一个药品, 令 $\overline{d_1}$ 为多个症状的平均向量距离, K 为症状个数, d_j 为单个症状与药品的向量距离, 那么对于一个药品的平均向量距离为: $\overline{d_1} = \sum_{j=1}^K d_j / K$ 。然后求出症状信息与每个药品的平均向量距离, 之后进行平均向量距离的排序, 将平均向量距离最短对应的药品作为目标药品。传统的文本匹配的方式对于症状的顺序不同, 搜索到的内容很可能不同, 且病人有多个症状时, 需要多次进行全量查询数据库, 效率低下, 而且查找到药品后, 怎么对药品进行排序也是一个很繁琐的工程。上述药品预测方法通过将文本匹配问题转换为了数学向量运算问题, 在预测速度上得到了极大的提升, 而且对于症状的顺序没有要求, 不同症状的顺序不一样也能得到一模一样的预测结果, 且通过向量距离的计算可以很容易地对搜索到的药品进行排序。故, 该药品预测方法相对于传统的文本匹配的方式, 不仅效率高, 而且灵活性、准确度和可操作性都大大得到了提高。

上述药品预测方法、装置、计算机设备及存储介质, 在获取到待预测药品的症状信息后, 确定与每个症状对应的症状向量, 然后根据每个症状的症状向

量，计算症状向量与药品数据库中每个药品的药品向量之间的向量距离，然后根据向量距离确定症状信息对应的目标药品。上述药品预测方法，通过将症状和药品之间的匹配关系转换为了向量运算，通过症状向量与药品向量之间的距离运算便可以快速找到目标药品，大大提高了查找的速度，即提高了预测药品的效率，且该查找方式有利于提高预测的准确度。

如图 3 所示，在一个实施例中，在确定与每个症状对应的症状向量之前，还包括：

步骤 210，获取每个药品对应的症状，得到症状训练样本集，症状训练样本集中包括多个症状训练样本。

其中，药品对应的症状的获取可以通过获取药品说明书中治疗的症状，整理药品说明书中治疗的症状，将症状作为一个个单词，用空格隔开。在一个实施例中，在获取到药品后，先对药品进行去重，将相同名称且治疗相同症状的药品去重，比如，有些规格、剂量不一样但是是相同的药品。在一个实施例中，将同一个药品的症状放在一起作为一个训练样本，即不同的药品对应不同的训练样本。由于一个药品治疗的症状往往只有几个，语料太少会导致训练效果不好，所以需要对称状数量进行拓展。在一个实施例中，可以通过反复复制其治疗的症状来拓展成新语料，比如，将头痛、发热、鼻塞三个症状复制两遍进行扩展，得到新的训练样本{头痛、发热、鼻塞、头痛、发热、鼻塞、头痛、发热、鼻塞}，从而扩大训练样本集。

步骤 212，将症状训练样本作为词向量模型的输入进行无监督训练，得到每个症状对应的症状向量。

其中，在得到症状训练样本后，将症状训练样本作为词向量模型的输入进行无监督训练，训练完成就可以得到每个症状对应的症状向量。词向量模型可以采用 word2vec 模型，词向量模型根据症状训练样本进行训练，能够将词共现多的词聚类，所以当症状与症状在一起出现的多时，他们得到的词向量在空间是很接近的。通过词向量模型可以得到每个症状对应的症状向量。

在一个实施例中，通过对每个药品的说明书中症状进行症状分词，然后采

用 word2vec 模型将症状训练成症状向量。Word2vec 分为 CBOW 和 Skip-gram 两种。CBOW 是根据上下文来预测当前词语的概率；Skip-gram 则刚好相反，是根据当前词语来预测上下文的概率。如图 4 所示，为 CBOW 和 Skip-gram 预测的原理示意图。 $w(t)$ 为文本中的某个词， $w(t-1)$ 和 $w(t+1)$ 分别为 $w(t)$ 在文本中的前一个词和后一个词。其中，症状向量的维数可以根据药品数量的多少来定，比如，可以采用 10 维，上下文窗口可以设置为 3。

步骤 214，根据药品对应的症状的症状向量计算得到药品对应的药品向量，将药品向量存储到药品数据库中。

其中，在计算得到每个症状对应的症状向量后，根据药品对应的症状就可以计算得到该药品对应的药品向量。在一个实施例中，可以将药品对应的多个症状向量进行平均得到平均向量，然后将平均向量作为该药品的药品向量。

每个药品的药效可以对应多个症状，如果用数学表示，药品可以在空间中用一个点表示，而药品治疗的症状则是分布在中心点周围的点。如图 5 所示，反应了不同药品与症状在三维空间分布降维到二维空间的可视化图。在图中，每个症状用一个点来表示，同一种类型的点对应的是同一个药品的症状。从图中可以看出，每个药品所治疗的症状在空间上分布是很接近的，不同药品的症状相互间离的比较远，图中药品的点未标出。

如图 6 所示，为一个实施例中，词向量模型训练的方法流程示意图。（1）先对药品进行去重，将相同名称且治疗相同症状的药品去重。（2）整理药品说明书中治疗的症状，将症状作为一个个的单词，用空格隔开。（3）对症状数量进行扩展，由于一个药品治疗的症状往往只有几个，语料太少训练效果不好，需要反复复制其治疗的症状拓展成新语料，扩大训练集。（4）采用 word2vec 模型将症状训练成症状向量，维数可以根据药品数量的多少来定。（5）存储症状向量，词向量模型反应了药品症状的分布。（6）利用每个药品所治疗的症状向量，求出每个药品的药品向量（各个症状的几何中心）。（7）存储每个药品的药品向量。

在一个实施例中，根据药品对应的症状的症状向量计算得到药品对应的药

品向量, 包括: 当药品对应多个症状时, 获取每个症状对应的症状向量; 计算多个症状向量的平均向量, 将平均向量作为相应药品的药品向量。

其中, 当药品对应多个症状时, 在得到每个症状对应的症状向量后, 根据每个症状的症状向量来确定药品的药品向量。可以将多个症状向量对应的几何中心作为相应药品的药品向量, 具体地, 将多个症状向量进行平均得到平均向量, 该平均向量为多个症状向量的几何中心对应的向量, 将该平均向量作为药品向量。

如图 7 所示, 在一个实施例中, 根据向量距离确定与症状信息对应的目标药品, 包括:

步骤 208A, 当症状信息中包括多个症状时, 根据每个症状向量与药品向量之间的向量距离计算得到平均向量距离。

其中, 症状信息中包括多个症状时, 分别计算每个症状向量与每个药品向量之间的向量距离, 根据多个症状向量与同一个药品向量之间的向量距离计算得到多个症状向量与药品向量之间的平均向量距离。

在一个实施例中, 如果是多个症状预测一个药品, 令 $\overline{d}_1$ 为多个症状的平均向量距离, K 为症状个数, d_j 为单个症状与药品的向量距离, 那么对于一个药品的平均向量距离为: $\overline{d}_1 = \sum_{i=1}^K d_j / K$ 。然后求出症状信息与每个药品的平均向量距离。

步骤 208B, 根据平均向量距离和症状个数计算症状信息对应的多个症状向量与药品向量之间的目标向量距离。

其中, 在预测药品时, 单个症状对药品预测的特征有限, 每多一个症状, 对药品的预测应该远大于单个症状, 所以为了提高预测的速度。需要根据症状个数设置出合理的权重, 以加快缩小向量距离的速度。在一个实施例中, 目标向量距离随着症状个数的增多而缩减。在一个具体的实施例中, 采用如下公式计算得到目标向量距离。当 $\overline{d}_1 \geq 1$ 时, $\overline{d} = \overline{d}_1^{1/K^4}$, 当 $0 < \overline{d}_1 < 1$ 时, $\overline{d} = \overline{d}_1^{K^4}$, 其中, $\overline{d}$ 表示加权完后的目标向量距离, $\overline{d}_1$ 为多个症状的平均向量距离, K 表示症状个数。通过上述公式, 每多一个症状, 与药品的空间距离将会成倍缩小。

步骤 208C, 根据多个症状向量与每个药品的药品向量之间的目标向量距离确定与症状信息对应的目标药品。

其中, 在计算得到多个症状向量与每个药品的药品向量之间的目标向量距离后, 将各个目标向量距离进行排序, 将最短的目标向量距离对应的药品作为
5 目标药品。

在一个实施例中, 在所述获取待预测药品的症状信息之前, 还包括: 获取问诊对话文本, 对所述问诊对话文本进行分词处理, 得到多个词语; 当在症状实体数据库中能够查找到所述词语时, 将所述词语作为所述症状信息中的症状。

其中, 问诊对话文本是指描述用户症状的文本。问诊对话文本可以是识别
10 用户语音得到的文本, 也可以是直接输入的文本。在得到问诊对话文本后, 对问诊对话文本进行分词处理, 得到多个词语。症状实体数据库中存储了各种症状的词语, 通过将得到的词语与症状实体数据库中词语进行匹配, 若能够在症状实体数据库中找到该词语, 说明该词语为描述症状的词语, 将该词语作为症状信息中的症状。在另一个实施例中, 在提取到问诊对话文本, 将问诊对话文
15 本输入症状实体识别模型, 通过该症状实体识别模型识别问诊对话文本得到相应的症状实体 (即描述症状的词语)。

在一个实施例中, 在所述获取问诊对话文本, 对所述问诊对话文本进行分词处理, 得到多个词语之后, 还包括: 获取词语映射关系表, 根据所述词语映射关系表获取与每个词语对应的目标词语; 所述当在症状实体数据库中能够查
20 找到所述词语时, 将所述词语作为所述症状信息中的症状, 包括: 当在症状实体数据库中能够查找到所述目标词语时, 则将所述目标词语作为所述症状信息中的症状。

其中, 由于同一症状的表述方式有多种, 比如, 描述头痛的症状的词语有
“脑壳痛”、“头疼”等口语化的词语。所以在对文本进行分词处理得到多个词
25 语后, 获取词语映射关系表, 在词语映射关系表中查找与每个词语对应的目标词语。词语映射关系表是指将口语化的症状词语转换为标准的症状词语 (目标词语)。当在症状实体数据库中能够查找到目标词语时, 则将目标词语作为症状

信息中的症状。比如，如果通过文本分词直接得到的词语为“脑壳痛”，那么词语映射关系表中记载了“脑壳痛”与“头痛”的映射关系，那么就可以得到目标词语“头痛”，之后将“头痛”作为症状信息中的症状。

在一个实施例中，在所述根据所述向量距离确定与所述症状信息对应的药品之后，还包括：获取所述目标药品对应的目标症状；将所述目标症状与所述症状信息中的症状进行比较，当所述目标症状包含有所述症状信息中的所有症状时，则判定所述目标药品安全。

其中，在预测得到目标药品后，为了确保药品预测的安全性，对预测到的药品进行症状检查，药品必须能够治疗用户所描述到的症状才能推荐。所以要获取目标药品对应的目标症状，然后将该目标症状与症状信息中的症状进行比较，只有当目标症状中包含有症状信息中的所有症状时，才确定预测得到的目标药品安全。在一个实施例中，在症状匹配时，可以分别获取目标症状的目标症状向量，和症状信息中症状的症状向量，根据目标症状向量和症状向量之间的向量距离来判断两者是否为同一个症状。比如，令症状 1 的症状向量为 v_1 、症状 2 的症状向量为 v_2 ，将向量距离设为 0 时，则说明症状 1 和症状 2 相同。

如图 8 所示为一个实施例中，药品预测的方法流程示意图。包括如下步骤：

(1) 获取问诊对话文本，对问诊对话文本进行分词。(2) 根据分词结果提取一个或多个症状（比如，“我有点头痛发烧”中的“头痛”和“发烧”）。(3) 根据训练好的词向量，获取每个症状对应的症状向量。(4) 计算症状与药品的向量距离，具体地，获取每个药品的药品向量，采用欧式距离计算症状与药品的向量距离，当多个症状预测一个药品时，则还需要计算多个症状与药品的平均向量距离。(5) 根据症状个数对向量距离进行加权，得到多个症状与药品的目标向量距离，根据目标向量距离确定预测得到的目标药品。(6) 对预测到的目标药品进行安全检查，目标药品必须能够治疗提出的所有症状。(7) 返回预测到的目标药品或者提示无药治疗这些症状。

如图 9 所示，在一个实施例中，提出了一种药品预测装置，所述装置包括：获取模块 902，用于获取待预测药品的症状信息，所述症状信息中包括至

少一个症状；

向量确定模块 904，用于确定与每个症状对应的症状向量；

计算模块 906，用于根据所述每个症状的症状向量，计算所述症状向量与药品数据库中的每个药品对应的药品向量之间的向量距离；

5 药品确定模块 908，用于根据所述向量距离确定与所述症状信息对应的目标药品。

如图 10 所示，在一个实施例中，上述药品预测装置还包括：

样本确定模块 910，用于获取每个药品对应的症状，得到症状训练样本集，所述症状训练样本集中包括多个症状训练样本；

10 输入输出模块 912，用于将所述症状训练样本作为词向量模型的输入进行无监督训练，得到每个症状对应的症状向量；

存储模块 914，用于根据药品对应的症状的症状向量计算得到所述药品对应的药品向量，将所述药品向量存储到所述药品数据库中。

15 在一个实施例中，所述存储模块还用于当所述药品对应多个症状时，获取每个症状对应的症状向量；计算多个所述症状向量的平均向量，将所述平均向量作为相应药品的药品向量。

20 在一个实施例中，药品确定模块还用于当所述症状信息中包括多个症状时，根据每个症状向量与药品向量之间的向量距离计算得平均向量距离；根据所述平均向量距离和症状个数计算所述症状信息对应的多个症状向量与所述药品向量之间的目标向量距离；根据所述多个症状向量与每个药品的药品向量之间的目标向量距离确定与所述症状信息对应的目标药品。

如图 11 所示，在一个实施例中，上述药品预测装置还包括：

文本获取模块 916，用于获取问诊对话文本，对所述问诊对话文本进行分词处理，得到多个词语。

25 症状确定模块 918，用于当在症状实体数据库中能够查找到所述词语时，将所述词语作为所述症状信息中的症状。

在一个实施例中，上述药品预测装置还包括：映射模块，用于获取词语映

射关系表，根据所述词语映射关系表获取与每个词语对应的目标词语；所述症状确定模块还用于当在症状实体数据库中能够查找到所述目标词语时，则将所述目标词语作为所述症状信息中的症状。

在一个实施例中，上述药品预测装置还包括：比较模块，用于获取所述目标药品对应的目标症状，将所述目标症状与所述症状信息中的症状进行比较，当所述目标症状包含有所述症状信息中的所有症状时，则判定所述目标药品安全。

图 12 示出了一个实施例中计算机设备的内部结构图。该计算机设备可以是终端，也可以是服务器。如图 12 所示，该计算机设备包括通过系统总线连接的处理器、存储器和网络接口。其中，存储器包括非易失性存储介质和内存储器。该计算机设备的非易失性存储介质存储有操作系统，还可存储有计算机程序，该计算机程序被处理器执行时，可使得处理器实现药品预测方法。该内存储器中也可储存有计算机程序，该计算机程序被处理器执行时，可使得处理器执行药品预测方法。网络接口用于与外界进行通信。本领域技术人员可以理解，图 12 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。

在一个实施例中，本申请提供的药品预测方法可以实现为一种计算机程序的形式，计算机程序可在如图 12 所示的计算机设备上运行。计算机设备的存储器中可存储组成该药品预测装置的各个程序模板。比如，获取模块 902、向量确定模块 904、计算模块 906 和药品确定模块 908。

一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行如下步骤：获取待预测药品的症状信息，所述症状信息中包括至少一个症状；确定与每个症状对应的症状向量；根据所述每个症状的症状向量，计算所述症状向量与药品数据库中的每个药品对应的药品向量之间的向量距离；根据所述向量距离确定与所述症状信息对应的目标药品。

一种计算机可读存储介质，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行如下步骤：获取待预测药品的症状信息，所述症状信息中包括至少一个症状；确定与每个症状对应的症状向量；根据所述每个症状的症状向量，计算所述症状向量与药品数据库中的每个药品对应的药品向量之间的向量距离；根据所述向量距离确定与所述症状信息对应的目标药品。

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，是可以通过计算机程序来指令相关的硬件来完成，所述的程序可存储于一非易失性计算机可读存储介质中，该程序在执行时，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本申请专利范围的限制。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

1、一种药品预测方法，其特征在于，所述方法包括：

获取待预测药品的症状信息，所述症状信息中包括至少一个症状；

确定与每个症状对应的症状向量；

5 根据所述每个症状的症状向量，计算所述症状向量与药品数据库中的每个药品对应的药品向量之间的向量距离；

根据所述向量距离确定与所述症状信息对应的目标药品。

2、根据权利要求1所述的方法，其特征在于，在所述确定与每个症状对应的症状向量之前，还包括：

10 获取每个药品对应的症状，得到症状训练样本集，所述症状训练样本集中包括多个症状训练样本；

将所述症状训练样本作为词向量模型的输入进行无监督训练，得到每个症状对应的症状向量；

15 根据药品对应的症状的症状向量计算得到所述药品对应的药品向量，将所述药品向量存储到所述药品数据库中。

3、根据权利要求2所述的方法，其特征在于，所述根据药品对应的症状的症状向量计算得到所述药品对应的药品向量，包括：

当所述药品对应多个症状时，获取每个症状对应的症状向量；

20 计算多个所述症状向量的平均向量，将所述平均向量作为相应药品的药品向量。

4、根据权利要求1所述的方法，其特征在于，所述根据所述向量距离确定与所述症状信息对应的目标药品，包括：

当所述症状信息中包括多个症状时，根据每个症状向量与药品向量之间的向量距离计算得平均向量距离；

25 根据所述平均向量距离和症状个数计算所述症状信息对应的多个症状向量与所述药品向量之间的目标向量距离；

根据所述多个症状向量与每个药品的药品向量之间的目标向量距离确定与

所述症状信息对应的目标药品。

5、根据权利要求1所述的方法，其特征在于，在所述获取待预测药品的症状信息之前，还包括：

获取问诊对话文本，对所述问诊对话文本进行分词处理，得到多个词语；

5 当在症状实体数据库中能够查找到所述词语时，将所述词语作为所述症状信息中的症状。

6、根据权利要求5所述的方法，其特征在于，在所述获取问诊对话文本，对所述问诊对话文本进行分词处理，得到多个词语之后，还包括：

10 获取词语映射关系表，根据所述词语映射关系表获取与每个词语对应的目标词语；

所述当在症状实体数据库中能够查找到所述词语时，将所述词语作为所述症状信息中的症状，包括：

当在症状实体数据库中能够查找到所述目标词语时，则将所述目标词语作为所述症状信息中的症状。

15 7、根据权利要求1所述的方法，其特征在于，在所述根据所述向量距离确定与所述症状信息对应的药品之后，还包括：

获取所述目标药品对应的目标症状；

将所述目标症状与所述症状信息中的症状进行比较，当所述目标症状包含有所述症状信息中的所有症状时，则判定所述目标药品安全。

20

8、一种药品预测装置，其特征在于，所述装置包括：

获取模块，用于获取待预测药品的症状信息，所述症状信息中包括至少一个症状；

向量确定模块，用于确定与每个症状对应的症状向量；

25 计算模块，用于根据所述每个症状的症状向量，计算所述症状向量与药品数据库中的每个药品对应的药品向量之间的向量距离；

药品确定模块，用于根据所述向量距离确定与所述症状信息对应的目标药

品。

9、一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行如权利要求 1 至 7 中任一项所述方法的步骤。

5 10、一种计算机可读存储介质，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行如权利要求 1 至 7 中任一项所述方法的步骤。

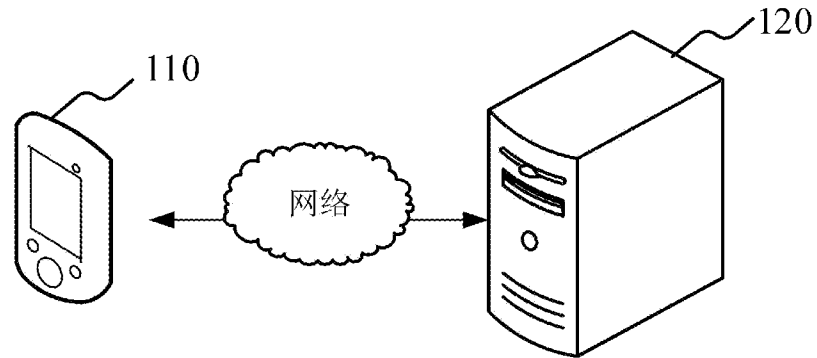


图 1

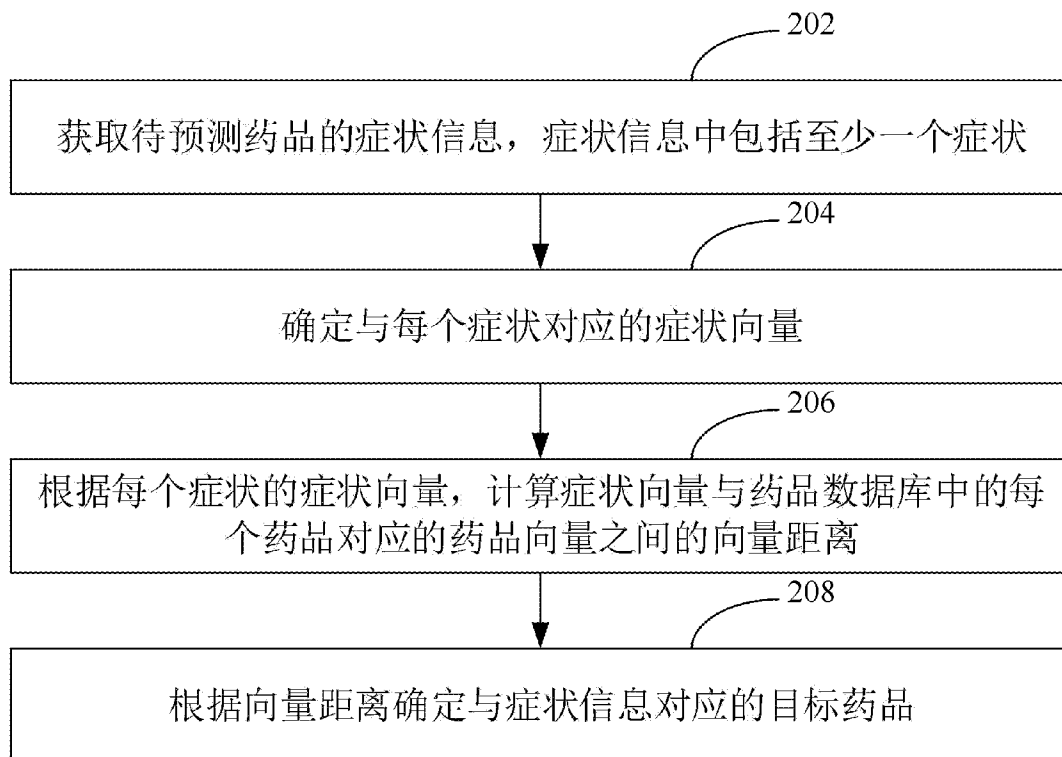


图 2

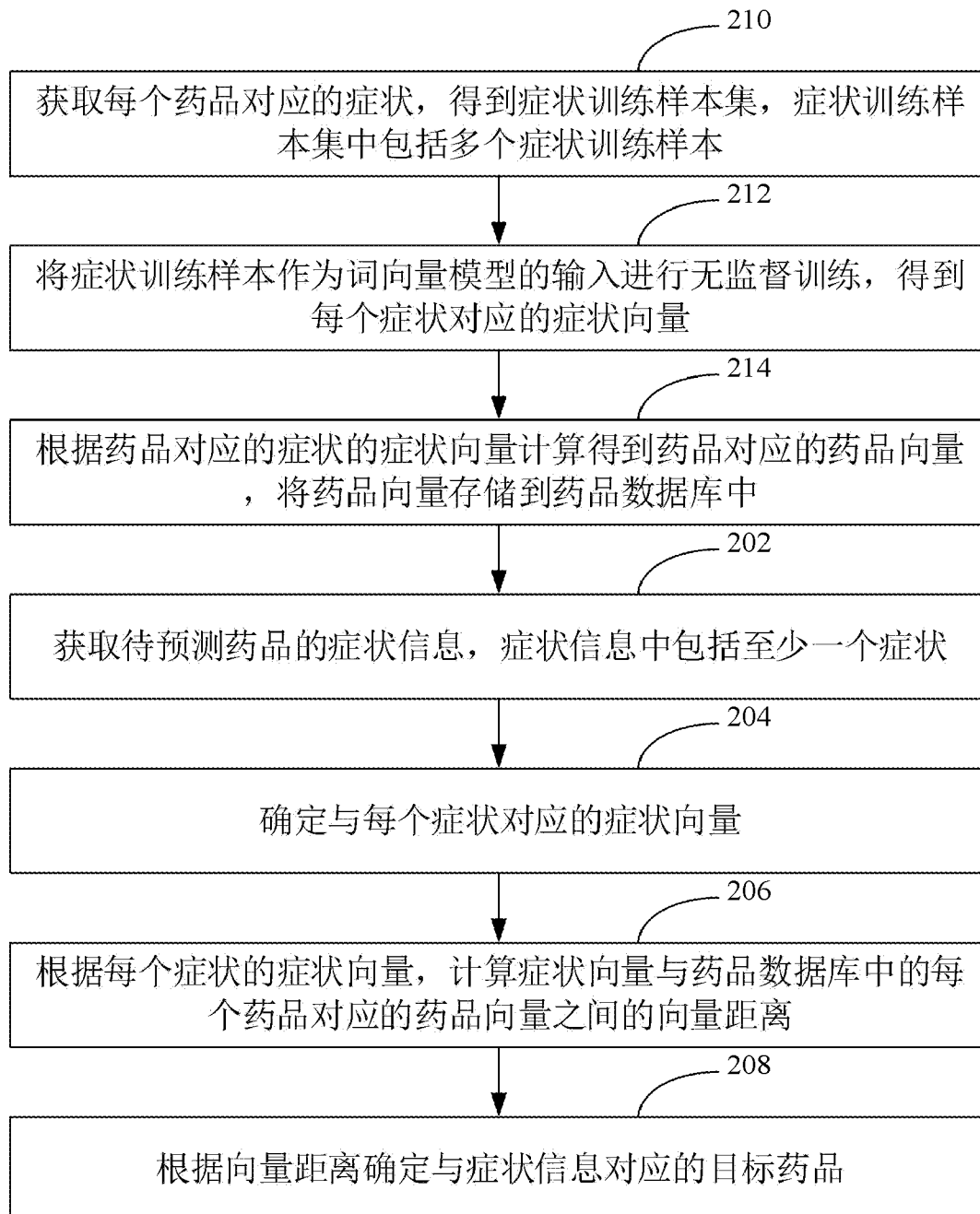


图 3

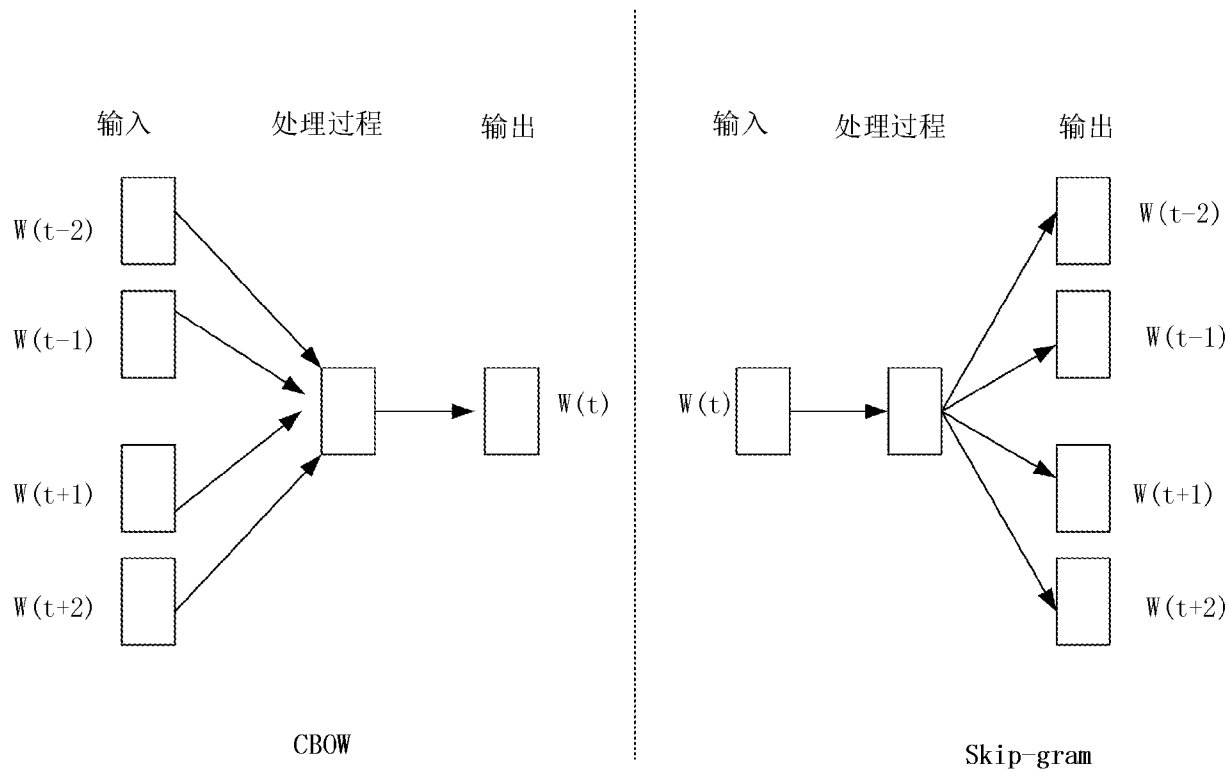


图 4

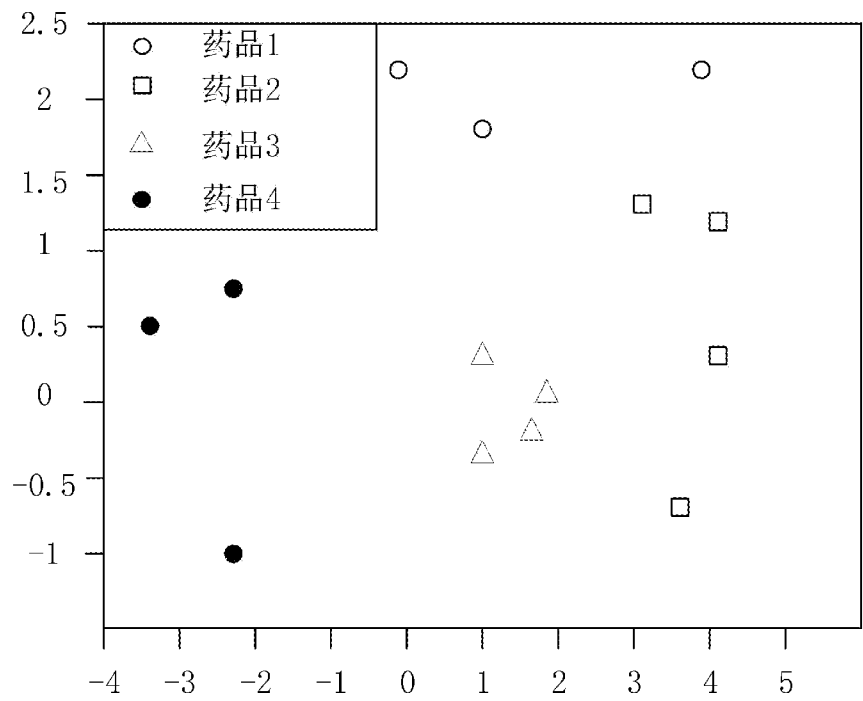


图 5

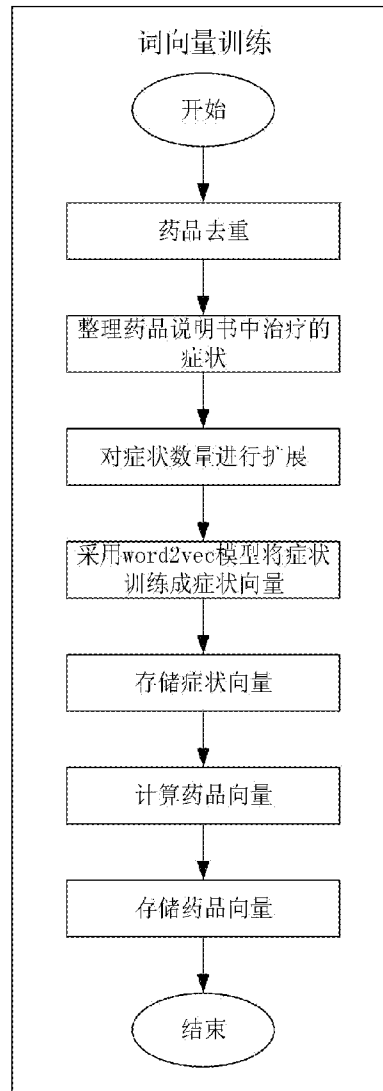


图 6

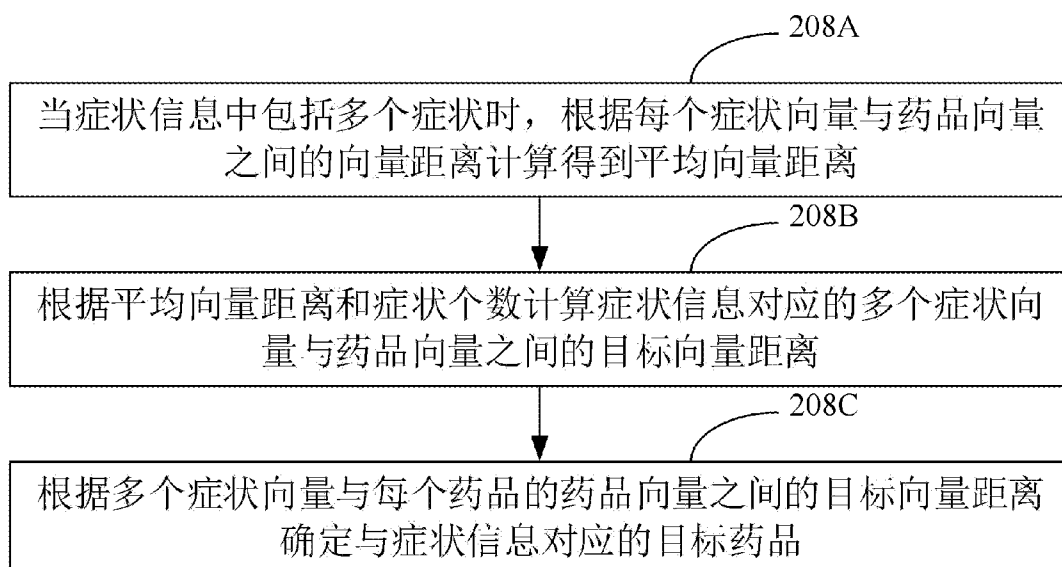


图 7

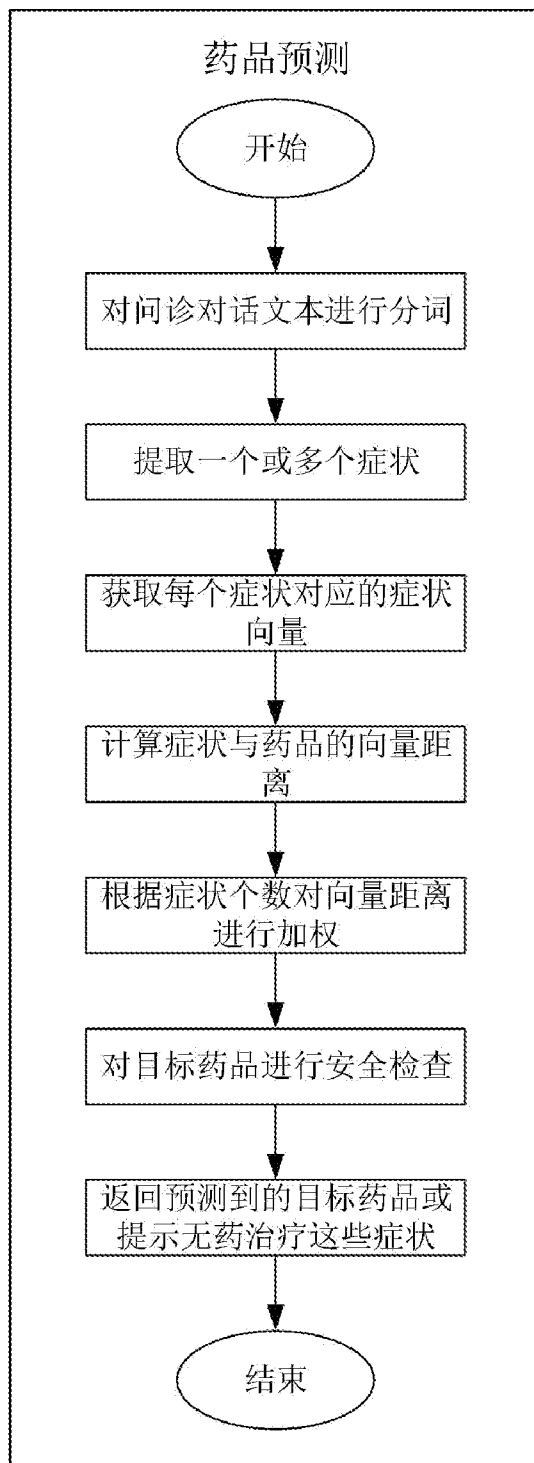


图 8

6/8

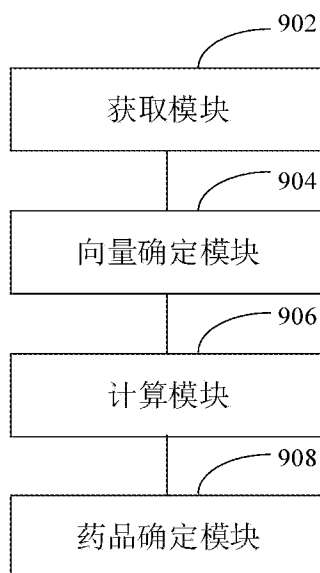


图 9

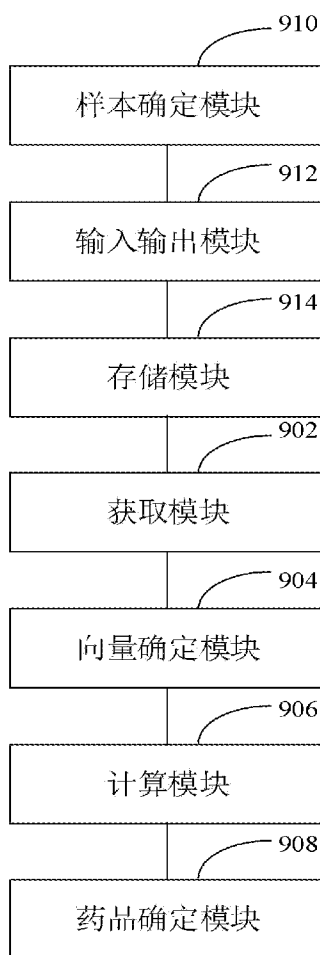


图 10

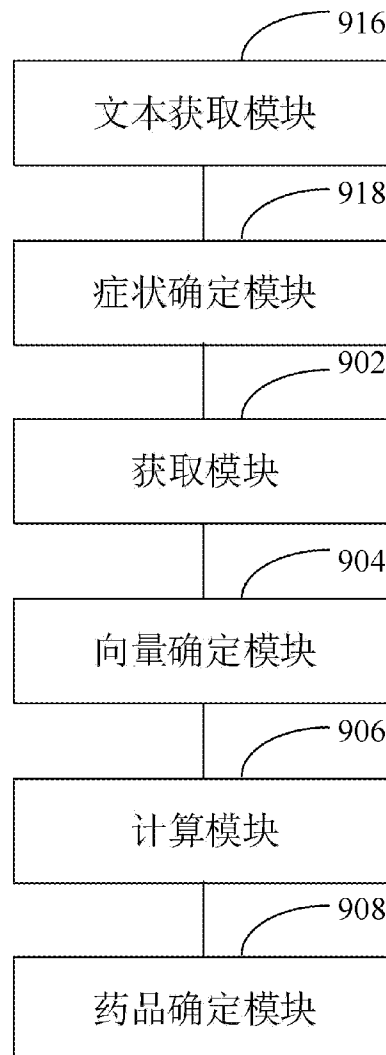


图 11

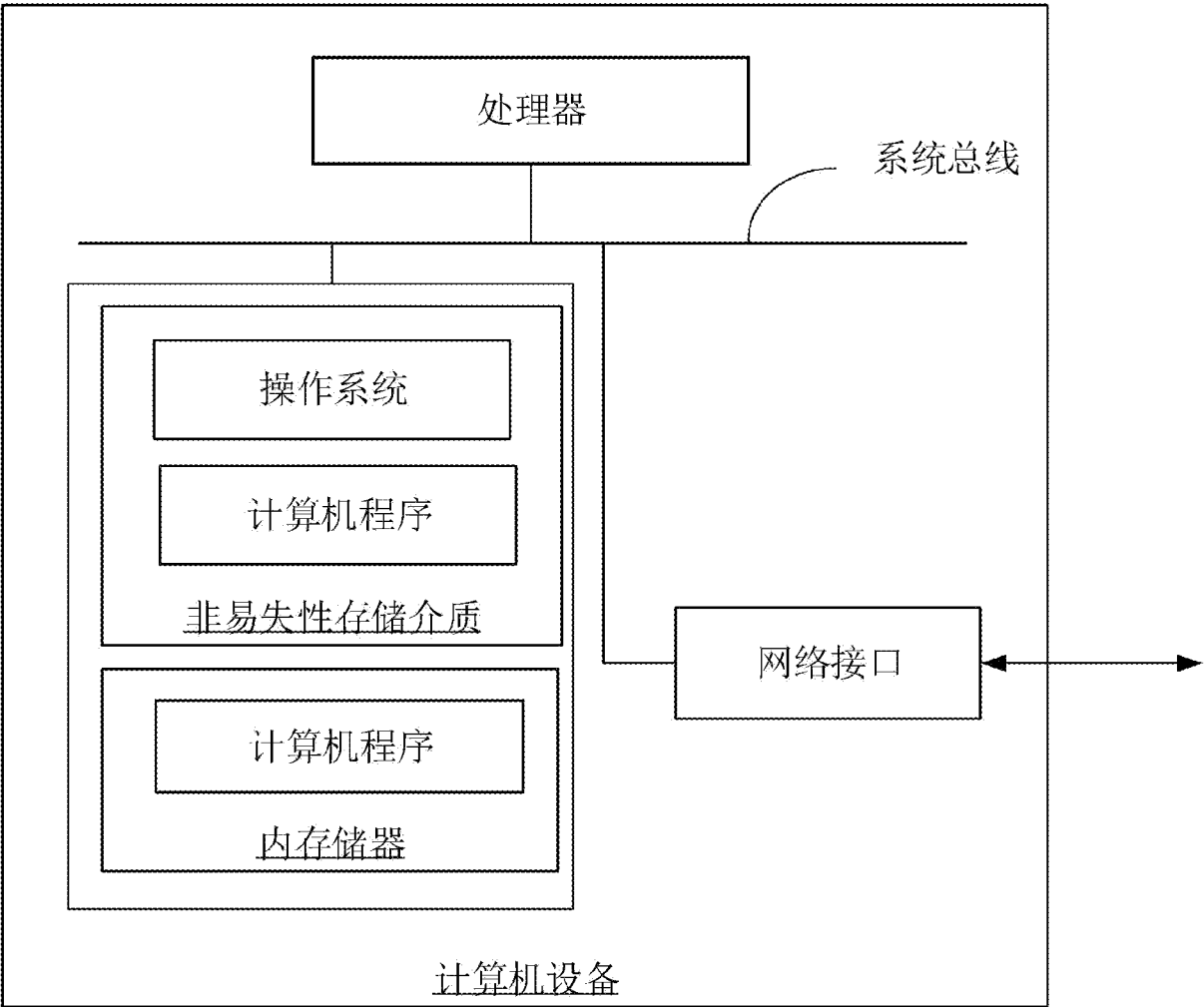


图 12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/123761

A. CLASSIFICATION OF SUBJECT MATTER

G16H 20/10(2018.01)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

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, WPI, EPODOC, CNKI, IEEE: 药, 症状, 向量, 训练, 距离, 推荐, 预测, medicine, symptom, vector, training, distance, commend, forecast

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107330288 A (DINGDANG (SHENZHEN) HEALTH ROBOT TECHNOLOGY CO., LTD.) 07 November 2017 (2017-11-07) claims 1-10, and description, paragraphs [0023]-[0062]	1-10
A	CN 107092797 A (GUANGDONG YIRONG ELECTRONIC COMMERCE CO., LTD. et al.) 25 August 2017 (2017-08-25) entire document	1-10
A	CN 107247868 A (DEEP THOUGHTS ARTIFICIAL INTELLIGENCE ROBOT TECHNOLOGY (BEIJING) CO., LTD.) 13 October 2017 (2017-10-13) entire document	1-10
A	CN 107591189 A (CSG SMART SCIENCE & TECHNOLOGY CO., LTD.) 16 January 2018 (2018-01-16) entire document	1-10
A	WO 2018006629 A1 (BEIJING SOGOU TECHNOLOGY DEVELOPMENT CO., LTD.) 11 January 2018 (2018-01-11) entire document	1-10



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

04 September 2019

Date of mailing of the international search report

30 September 2019

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/123761

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107330288	A	07 November 2017	None			
CN	107092797	A	25 August 2017	None			
CN	107247868	A	13 October 2017	None			
CN	107591189	A	16 January 2018	None			
WO	2018006629	A1	11 January 2018	CN	107590146	A	16 January 2018

国际检索报告

国际申请号

PCT/CN2018/123761

A. 主题的分类

G16H 20/10(2018.01)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

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, WPI, EPDOC, CNKI, IEEE: 药, 症状, 向量, 训练, 距离, 推荐, 预测, medicine, symptom, vector, training, distance, commend, forecast

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
X	CN 107330288 A (叮当深圳健康机器人科技有限公司) 2017年 11月 7日 (2017 - 11 - 07) 权利要求1-10、说明书第[0023]-[0062]段	1-10
A	CN 107092797 A (广东亿荣电子商务有限公司 等) 2017年 8月 25日 (2017 - 08 - 25) 全文	1-10
A	CN 107247868 A (深思考人工智能机器人科技北京有限公司) 2017年 10月 13日 (2017 - 10 - 13) 全文	1-10
A	CN 107591189 A (科大智能科技股份有限公司) 2018年 1月 16日 (2018 - 01 - 16) 全文	1-10
A	WO 2018006629 A1 (北京搜狗科技发展有限公司) 2018年 1月 11日 (2018 - 01 - 11) 全文	1-10

☐ 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 9月 4日

国际检索报告邮寄日期

2019年 9月 30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

胡贝贝

电话号码 86-(10)-53961415

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/123761

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	107330288	A	2017年 11月 7日	无			
CN	107092797	A	2017年 8月 25日	无			
CN	107247868	A	2017年 10月 13日	无			
CN	107591189	A	2018年 1月 16日	无			
WO	2018006629	A1	2018年 1月 11日	CN	107590146	A	2018年 1月 16日