



(12) 发明专利

(10) 授权公告号 CN 115620886 B

(45) 授权公告日 2023.04.28

(21) 申请号 202211628930.2

(22) 申请日 2022.12.19

(65) 同一申请的已公布的文献号
申请公布号 CN 115620886 A

(43) 申请公布日 2023.01.17

(73) 专利权人 北京大学第三医院(北京大学第
三临床医学院)

地址 100191 北京市海淀区花园北路49号

专利权人 北京京东世纪贸易有限公司

(72) 发明人 闫盈盈 徐晓涵 翟所迪 杨帅
张亚 周谦

(74) 专利代理机构 北京中和立达知识产权代理
有限公司 11756

专利代理师 张攀

(51) Int.Cl.

G16H 40/20 (2018.01)

G06F 16/36 (2019.01)

(56) 对比文件

CN 110782998 A, 2020.02.11

CN 111191020 A, 2020.05.22

审查员 黄莹

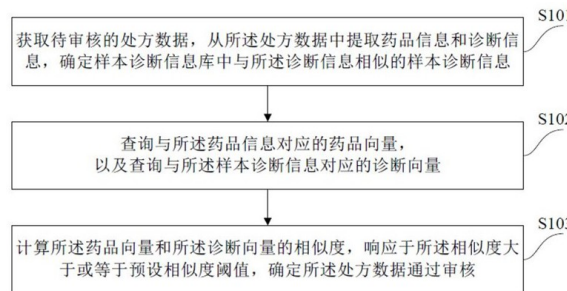
权利要求书3页 说明书10页 附图5页

(54) 发明名称

一种数据审核方法和装置

(57) 摘要

本发明公开了一种数据审核方法和装置,涉及智慧医疗技术领域。该方法的一具体实施方式包括:获取待审核的处方数据,从所述处方数据中提取药品信息和诊断信息,确定样本诊断信息库中与所述诊断信息相似的样本诊断信息;查询与所述药品信息对应的药品向量,以及查询与所述样本诊断信息对应的诊断向量;计算所述药品向量和所述诊断向量的相似度,响应于所述相似度大于或等于预设相似度阈值,确定所述处方数据通过审核。该实施方式基于知识图谱挖掘药品信息与药品信息、药品信息与诊断信息、诊断信息与诊断信息之间的关系,能够解决处方数据中诊断描述多样化的问题。



1. 一种数据审核方法,其特征在于,包括:

获取待审核的处方数据,从所述处方数据中提取药品信息和诊断信息,确定样本诊断信息库中与所述诊断信息相似的样本诊断信息;

查询与所述药品信息对应的药品向量,以及查询与所述样本诊断信息对应的诊断向量;其中,药品向量和诊断向量通过元路径训练得到,元路径表示知识图谱中从药品信息到诊断信息之间的路径;

计算所述药品向量和所述诊断向量的相似度,响应于所述相似度大于或等于预设相似度阈值,确定所述处方数据通过审核;

所述确定样本诊断信息库中与所述诊断信息相似的样本诊断信息,包括:

分别计算所述诊断信息和样本诊断信息库中每个样本诊断信息的相似度,将相似度最高的样本诊断信息,作为与所述诊断信息相似的目标样本诊断信息;

调用诊断属性提取模型,从所述诊断信息中提取第四属性、第五属性和第六属性;

确定所述样本诊断信息库中与所述诊断信息的第四属性相似的第一样本诊断信息集合、与第五属性相似的第二样本诊断信息集合、以及与第六属性相似的第三样本诊断信息集合;

对所述第一样本诊断信息集合、所述第二样本诊断信息集合、所述第三样本诊断信息集合做交集,响应于交集集合中的样本诊断信息数量仅为一个,得到与所述诊断信息相似的目标样本诊断信息;或

响应于交集集合中的样本诊断信息数量为多个,分别计算所述诊断信息和交集集合中每个样本诊断信息的相似度,将相似度最高的样本诊断信息,作为与所述诊断信息相似的目标样本诊断信息;

所述的第四属性为病理过程属性;

所述的第五属性为发病部位属性;

所述的第六属性为子症状属性。

2. 根据权利要求1所述的方法,其特征在于,在获取所述处方数据之前,还包括:

接收输入的样本药品诊断配对关系;其中,样本药品诊断配对关系包括匹配的样本药品信息和样本诊断信息;

查询与所述样本药品信息对应的药品说明书信息,从所述药品说明书信息中提取药品特征信息,以及从所述样本诊断信息中提取诊断特征信息;

基于所述药品特征信息和所述诊断特征信息构建知识图谱,使用所述知识图谱对机器学习模型进行训练,得到词向量模型,以基于所述词向量模型得到药品向量和诊断向量。

3. 根据权利要求2所述的方法,其特征在于,所述从所述药品说明书信息中提取药品特征信息,包括:调用药品属性提取模型,从所述药品说明书信息中提取药品通用名、第一属性、第二属性和第三属性,以构建药品特征信息;

所述从所述样本诊断信息中提取诊断特征信息,包括:调用诊断属性提取模型,从所述样本诊断信息中提取诊断名称、第四属性、第五属性和第六属性,查询与所述第四属性对应的第七属性,以构建诊断特征信息;

所述的第一属性为功效属性;

所述的第二属性为治疗部位属性;

所述的第三属性为治疗适应症属性；

所述的第七属性为治法属性。

4. 根据权利要求3所述的方法,其特征在于,所述基于所述药品特征信息和所述诊断特征信息构建知识图谱,包括:

基于药品通用名、第一属性、第七属性、第四属性、诊断名称,构建第一元路径;以及

基于药品通用名、第二属性、第五属性、诊断名称,构建第二元路径;以及

基于药品通用名、第三属性、第六属性、诊断名称,构建第三元路径;

基于所述第一元路径、所述第二元路径和所述第三元路径构建知识图谱。

5. 一种数据审核装置,其特征在于,包括:

提取模块,用于获取待审核的处方数据,从所述处方数据中提取药品信息和诊断信息,确定样本诊断信息库中与所述诊断信息相似的样本诊断信息;

查询模块,用于查询与所述药品信息对应的药品向量,以及查询与所述样本诊断信息对应的诊断向量;其中,药品向量和诊断向量通过元路径训练得到,元路径表示知识图谱中从药品信息到诊断信息之间的路径;

计算模块,用于计算所述药品向量和所述诊断向量的相似度,响应于所述相似度大于或等于预设相似度阈值,确定所述处方数据通过审核;

所述确定样本诊断信息库中与所述诊断信息相似的样本诊断信息,包括:

分别计算所述诊断信息和样本诊断信息库中每个样本诊断信息的相似度,将相似度最高的样本诊断信息,作为与所述诊断信息相似的目标样本诊断信息;

所述确定样本诊断信息库中与所述诊断信息相似的样本诊断信息,还包括:

调用诊断属性提取模型,从所述诊断信息中提取第四属性、第五属性和第六属性;

确定所述样本诊断信息库中与所述诊断信息的第四属性相似的第一样本诊断信息集合、与第五属性相似的第二样本诊断信息集合、以及与第六属性相似的第三样本诊断信息集合;

对所述第一样本诊断信息集合、所述第二样本诊断信息集合、所述第三样本诊断信息集合做交集,响应于交集集合中的样本诊断信息数量仅为一个,得到与所述诊断信息相似的目标样本诊断信息;或

响应于交集集合中的样本诊断信息数量为多个,分别计算所述诊断信息和交集集合中每个样本诊断信息的相似度,将相似度最高的样本诊断信息,作为与所述诊断信息相似的目标样本诊断信息;

所述的第四属性为病理过程属性;

所述的第五属性为发病部位属性;

所述的第六属性为子症状属性。

6. 一种电子设备,其特征在于,包括:

一个或多个处理器;

存储装置,用于存储一个或多个程序,

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如权利要求1-4中任一项所述的方法。

7. 一种计算机可读介质,其上存储有计算机程序,其特征在于,所述程序被处理器执行

时实现如权利要求1-4中任一项所述的方法。

一种数据审核方法和装置

技术领域

[0001] 本发明涉及智慧医疗技术领域,尤其涉及一种数据审核方法和装置。

背景技术

[0002] 由于医疗资源分配的不均,导致部分医院药师的专业领域知识缺乏,对医院开具的处方审核不够完善及准确。为解决这一问题设置了处方审核系统,以基于客观的药学规则审查处方中适应症部分。

[0003] 目前主要基于药品说明书中的适应症描述进行审查,但多数处方中诊断的描述与药品说明书中的适应症描述不同,且随着医疗技术的发展,出现很多超出药品说明书用药的使用方式。另外,目前适应症审查功能需基于药师手工维护,当处方中涉及全新的诊断或者是药品说明书适应症的同义词时,现有处方审核系统无法有效解决上述技术问题。

发明内容

[0004] 有鉴于此,本发明提供一种数据审核方法和装置,至少能够解决现有技术中由于诊断描述的多样性,导致准确地对方剂适应症进行审查异常困难的问题。

[0005] 为实现上述目的,根据本发明实施例的一个方面,提供了一种数据审核方法,包括:

[0006] 获取待审核的处方数据,从所述处方数据中提取药品信息和诊断信息,确定样本诊断信息库中与所述诊断信息相似的样本诊断信息;

[0007] 查询与所述药品信息对应的药品向量,以及查询与所述样本诊断信息对应的诊断向量;其中,药品向量和诊断向量通过元路径训练得到,元路径表示知识图谱中从药品信息到诊断信息之间的路径;

[0008] 计算所述药品向量和所述诊断向量的相似度,响应于所述相似度大于或等于预设相似度阈值,确定所述处方数据通过审核。

[0009] 可选的,在获取所述处方数据之前,还包括:

[0010] 接收输入的样本药品诊断配对关系;其中,样本药品诊断配对关系包括匹配的样本药品信息和样本诊断信息;

[0011] 查询与所述样本药品信息对应的药品说明书信息,从所述药品说明书信息中提取药品特征信息,以及从所述样本诊断信息中提取诊断特征信息;

[0012] 基于所述药品特征信息和所述诊断特征信息构建知识图谱,使用所述知识图谱对机器学习模型进行训练,得到词向量模型,以基于所述词向量模型得到药品向量和诊断向量。

[0013] 可选的,所述从所述药品说明书信息中提取药品特征信息,包括:调用药品属性提取模型,从所述药品说明书信息中提取药品通用名、第一属性、第二属性和第三属性,以构建药品特征信息;

[0014] 所述从所述样本诊断信息中提取诊断特征信息,包括:调用诊断属性提取模型,从

所述样本诊断信息中提取诊断名称、第四属性、第五属性和第六属性，查询与所述第四属性对应的第七属性，以构建诊断特征信息。

[0015] 可选的，所述基于所述药品特征信息和所述诊断特征信息构建知识图谱，包括：

[0016] 基于药品通用名、第一属性、第七属性、第四属性、诊断名称，构建第一元路径；以及

[0017] 基于药品通用名、第二属性、第五属性、诊断名称，构建第二元路径；以及

[0018] 基于药品通用名、第三属性、第六属性、诊断名称，构建第三元路径；

[0019] 基于所述第一元路径、所述第二元路径和所述第三元路径构建知识图谱。

[0020] 可选的，所述确定样本诊断信息库中与所述诊断信息相似的样本诊断信息，包括：

[0021] 分别计算所述诊断信息和样本诊断信息库中每个样本诊断信息的相似度，将相似度最高的样本诊断信息，作为与所述诊断信息相似的目标样本诊断信息。

[0022] 可选的，所述确定样本诊断信息库中与所述诊断信息相似的样本诊断信息，包括：

[0023] 调用诊断属性提取模型，从所述诊断信息中提取第四属性、第五属性和第六属性；

[0024] 确定所述样本诊断信息库中与所述诊断信息的第四属性相似的第一样本诊断信息集合、与第五属性相似的第二样本诊断信息集合、以及与第六属性相似的第三样本诊断信息集合；

[0025] 对所述第一样本诊断信息集合、所述第二样本诊断信息集合、所述第三样本诊断信息集合做交集，响应于交集集合中的样本诊断信息数量仅为一个，得到与所述诊断信息相似的目标样本诊断信息；或

[0026] 响应于交集集合中的样本诊断信息数量为多个，分别计算所述诊断信息和交集集合中每个样本诊断信息的相似度，将相似度最高的样本诊断信息，作为与所述诊断信息相似的目标样本诊断信息。

[0027] 为实现上述目的，根据本发明实施例的另一方面，提供了一种数据审核装置，包括：

[0028] 提取模块，用于获取待审核的处方数据，从所述处方数据中提取药品信息和诊断信息，确定样本诊断信息库中与所述诊断信息相似的样本诊断信息；

[0029] 查询模块，用于查询与所述药品信息对应的药品向量，以及查询与所述样本诊断信息对应的诊断向量；其中，药品向量和诊断向量通过元路径训练得到，元路径表示知识图谱中从药品信息到诊断信息之间的路径；

[0030] 计算模块，用于计算所述药品向量和所述诊断向量的相似度，响应于所述相似度大于或等于预设相似度阈值，确定所述处方数据通过审核。

[0031] 可选的，还包括训练模块，用于：

[0032] 接收输入的样本药品诊断配对关系；其中，样本药品诊断配对关系包括匹配的样本药品信息和样本诊断信息；

[0033] 查询与所述样本药品信息对应的药品说明书信息，从所述药品说明书信息中提取药品特征信息，以及从所述样本诊断信息中提取诊断特征信息；

[0034] 基于所述药品特征信息和所述诊断特征信息构建知识图谱，使用所述知识图谱对机器学习模型进行训练，得到词向量模型，以基于所述词向量模型得到药品向量和诊断向量。

[0035] 可选的,所述训练模块,用于:调用药品属性提取模型,从所述药品说明书信息中提取药品通用名、第一属性、第二属性和第三属性,以构建药品特征信息;

[0036] 所述训练模块,用于:调用诊断属性提取模型,从所述样本诊断信息中提取诊断名称、第四属性、第五属性和第六属性,查询与所述第四属性对应的第七属性,以构建诊断特征信息。

[0037] 可选的,所述训练模块,用于:

[0038] 基于药品通用名、第一属性、第七属性、第四属性、诊断名称,构建第一元路径;以及

[0039] 基于药品通用名、第二属性、第五属性、诊断名称,构建第二元路径;以及

[0040] 基于药品通用名、第三属性、第六属性、诊断名称,构建第三元路径;

[0041] 基于所述第一元路径、所述第二元路径和所述第三元路径构建知识图谱。

[0042] 可选的,所述提取模块,用于:

[0043] 分别计算所述诊断信息和样本诊断信息库中每个样本诊断信息的相似度,将相似度最高的样本诊断信息,作为与所述诊断信息相似的目标样本诊断信息。

[0044] 可选的,所述提取模块,用于:

[0045] 调用诊断属性提取模型,从所述诊断信息中提取第四属性、第五属性和第六属性;

[0046] 确定所述样本诊断信息库中与所述诊断信息的第四属性相似的第一样本诊断信息集合、与第五属性相似的第二样本诊断信息集合、以及与第六属性相似的第三样本诊断信息集合;

[0047] 对所述第一样本诊断信息集合、所述第二样本诊断信息集合、所述第三样本诊断信息集合做交集,响应于交集集合中的样本诊断信息数量仅为一个,得到与所述诊断信息相似的目标样本诊断信息;或

[0048] 响应于交集集合中的样本诊断信息数量为多个,分别计算所述诊断信息和交集集合中每个样本诊断信息的相似度,将相似度最高的样本诊断信息,作为与所述诊断信息相似的目标样本诊断信息。

[0049] 为实现上述目的,根据本发明实施例的再一方面,提供了一种数据审核电子设备。

[0050] 本发明实施例的电子设备包括:一个或多个处理器;存储装置,用于存储一个或多个程序,当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现上述任一所述的数据审核方法。

[0051] 为实现上述目的,根据本发明实施例的再一方面,提供了一种计算机可读介质,其上存储有计算机程序,所述程序被处理器执行时实现上述任一所述的数据审核方法。

[0052] 根据本发明所述提供的方案,上述发明具有如下优点或有益效果:将基于知识图谱的推荐算法引入处方适应证审查场景,挖掘药品通用名与诊断之间的关系,最终构建药品通用名向量和诊断向量,后续使用可以基于药品通用名的唯一性直接查询到相应向量,而对于诊断信息,则通过属性相似度计算、文本相似度计算、关键词相似度计算方式,确定与其相似的样本诊断信息、查询该样本诊断信息的诊断向量,以提升本次使用的诊断向量查询的高效性和准确性,整体摒弃现有应用时仍依赖模型计算向量的方式,能够解决诊断描述多样化的问题。

[0053] 上述的非惯用的可选方式所具有的进一步效果将在下文中结合具体实施方式加

以说明。

附图说明

- [0054] 附图用于更好地理解本发明,不构成对本发明的不当限定。其中:
- [0055] 图1是根据本发明实施例的一种数据审核方法的主要流程示意图;
- [0056] 图2是类比知识图谱的推荐算法示意图;
- [0057] 图3是根据本发明实施例的一种可选的数据审核方法的流程示意图;
- [0058] 图4(a)是药品知识图谱和疾病知识图谱的结构图;
- [0059] 图4(b)是知识图谱样例示意图;
- [0060] 图5是根据本发明实施例的另一种可选的数据审核方法的流程示意图;
- [0061] 图6是根据本发明实施例的一种数据审核装置的主要模块示意图;
- [0062] 图7是本发明实施例可以应用于其中的示例性系统架构图;
- [0063] 图8是适于用来实现本发明实施例的移动设备或服务器的计算机系统的结构示意图。

具体实施方式

[0064] 以下结合附图对本发明的示范性实施例做出说明,其中包括本发明实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本发明的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0065] 需要指出的是,在不冲突的情况下,本发明中的实施例以及实施例中的特征可以互相组合。本申请技术方案中对数据(如用户个人信息)的获取、存储、使用、处理等均符合国家法律法规的相关规定,且不违背公序良俗。

[0066] 参见图1,示出的是本发明实施例提供的一种数据审核方法的主要流程图,包括如下步骤:

[0067] S101:获取待审核的处方数据,从所述处方数据中提取药品信息和诊断信息,确定样本诊断信息库中与所述诊断信息相似的样本诊断信息;

[0068] S102:查询与所述药品信息对应的药品向量,以及查询与所述样本诊断信息对应的诊断向量;其中,药品向量和诊断向量通过元路径训练得到,元路径表示知识图谱中从药品信息到诊断信息之间的路径;

[0069] S103:计算所述药品向量和所述诊断向量的相似度,响应于所述相似度大于或等于预设相似度阈值,确定所述处方数据通过审核。

[0070] 上述实施方式中,药品说明书适应症的同义词,比如脑卒中和中风,说明书里可能描述的是脑卒中,但诊断中可能写的是中风,普通的字符串匹配无法解决这种问题,需要维护同义词表或者依靠知识图谱的推荐算法。

[0071] 处方适应症审核的目的在于判断药品与处方中的诊断是否匹配。参见图2所示,若将药品比作消费者,那么药品的属性就可以类比为消费者的喜好,若将诊断比作物品,那么疾病的相关属性就可以类比为物品的属性。因而,处方适应症审核可以转化为物品推荐问题,亦即判断消费者对推荐的物品是否感兴趣。在物品推荐的应用领域中,推荐算法中的物

品可以指代视频、文章、商品、新闻等一切内容对象,也可用项目指代。

[0072] 对于步骤S101,用户的医疗处方数据,为在用户就诊过程中所采集的数据,具体包括药品信息和诊断信息。诊断信息是用于表征诊断结果的数据,比如感冒、发热、咽痛,至少包括问诊信息、诊断结果和用药医嘱,而用药医嘱是医师根据临床场景对药物的调整使用备注,基于该备注可以使用药物适用症与当前诊断结果匹配度不是很高的药品,从而提高了药品的适应性使用,提高了医疗资源的利用率。

[0073] 本方案中的药品信息主要指的是药品通用名。药品通用名,指列入国家药品标准的药品名称,在全世界都可通用的名称,如阿司匹林,任何药品说明书上都应标注通用名,具有强制性和约束性。商品名是药品生产厂商自己确定,经药品监督管理部门核准的产品名称,具有专有性质,不得仿用。一个药品通用名下由于生产厂家的不同,可有多个商品名称,如头孢哌酮钠说明书标注:【通用名称】头孢哌酮钠;【商品名称】先抗;【别名】先锋必素、头孢氧哌唑、先锋必、先锋哌酮、氧哌羟苯唑、头孢菌素钠、先锋哌唑酮、先锋松、头孢氧哌羟苯唑、氧哌嗪头孢。处方中通常会药品有通用名,所以不涉及商用名的概念。

[0074] 对于步骤S102~S103,本方案预先构建药品通用名向量表和诊断向量表,药品通用名向量表以存储药品通用名向量,诊断向量表用以存储诊断向量,且该过程为离线操作,参见后续图3、图4(a)和图4(b)所示。

[0075] 从药品通用名向量表中,查询与本次药品通用名对应的药品通用名向量表示,以及从诊断向量表中,查询与本次诊断信息对应的诊断向量表示。通过计算两个向量的相似度,在一些可行的实施方式中,相似度的计算可以基于皮尔逊相关系数(Pearson Correlation Coefficient)、欧几里德距离(Euclidean Distance)等,本方案优选余弦相似度。当所得相似度如果超过一定的阈值(例如0.7),可认为用户医疗处方数据中的药品信息和诊断信息匹配,该处方通过审核,否则不通过。本实施方式为在线操作,如前述获取用户的医疗处方数据,输入的是药品通用名和诊断信息,输出则是处方数据是否通过审核。

[0076] 上述实施例所提供的方法,摒弃现有应用时仍依赖模型计算向量的方式,本方案预先设置药品通用名向量和诊断向量,以基于药品通用名的唯一性直接查询到向量,诊断向量通过相似度计算方式查询,解决了诊断描述多样性问题,进一步提升对处方数据审核的效率和准确率。

[0077] 参见图3,示出了根据本发明实施例的一种可选的数据审核方法流程示意图,包括如下步骤:

[0078] S301:接收输入的样本药品诊断配对关系;其中,样本药品诊断配对关系包括匹配的样本药品信息和样本诊断信息;

[0079] S302:查询与所述样本药品信息对应的药品说明书信息,从所述药品说明书信息中提取药品特征信息,以及从所述样本诊断信息中提取诊断特征信息;

[0080] S303:基于所述药品特征信息和所述诊断特征信息构建知识图谱,使用所述知识图谱对机器学习模型进行训练,得到词向量模型,以基于所述词向量模型得到药品向量和诊断向量。

[0081] 上述实施方式中,对于步骤S301~S302,本实施方式在训练机器学习模型之前,需要根据样本药品信息和样本诊断信息,获取结构化的药品说明书信息和诊断信息,样本药品诊断配对关系可以从药师审核通过的处方获取得到。

[0082] 基于药品通用名调取出对应的药品使用说明书。结构化的药品说明书信息包括药品通用名、功效属性(即第一属性)、治疗部位属性(即第二属性)和治疗适应症属性(即第三属性),可通过药品属性提取模型从药品说明书中获得。

[0083] 结构化的诊断信息包括病理过程属性(即第四属性)、发病部位属性(即第五属性)和子症状属性(即第六属性),可通过诊断属性提取模型获得。预先设置治法属性表(如对疼痛的治法为止疼),可以从该表中查询对该病例过程属性的治法属性(即第七属性)。

[0084] 以药品通用名为根痛平颗粒、诊断名称为左肩周炎为例,所得药品特征信息包括:药品通用名为根痛平颗粒、功效属性为止痛、治疗部位属性为肩、治疗适应症属性为肩颈疼痛,所得诊断特征信息包括:诊断名称为左肩周炎、病理过程属性为疼痛、发病部位属性为肩、子症状属性为肩颈疼痛,查询到对疼痛的治法属性为止痛。

[0085] 对于步骤S303,本方案在已构建好的知识图谱上训练药品通用名和诊断的向量,知识图谱的具体构建流程不在本方案的讨论范围内,在此不做详细阐述。使用的推荐算法优选基于元路径的推荐算法,以进一步提升模型的可解释性。元路径也可以认为是一组文本内容,在自然语言处理领域,统计语言模型(即机器学习模型)是所有自然语言处理技术的基础,被广泛应用于文本处理、语音识别、机器翻译、分词、词性标注和信息检索等任务。简单来说,统计语言模型是用来计算一个句子的概率的概率模型,通常基于一个语料库来构建。基于神经网络来构建统计语言模型时,会将文本转换为数值张量,也即词向量化。

[0086] 参见图4(a)所示,本方案构建的知识图谱包括三条元路径,分别为:1)药品通用名-功效属性-治法属性-病理过程属性-诊断名称;2)药品通用名-治疗部位属性-发病部位属性-诊断名称;3)药品通用名-治疗适应症属性-子症状属性-诊断名称。参见图4(b)所示,基于根痛平颗粒、止痛、止痛、疼痛、左肩周炎构建第一元路径,基于根痛平颗粒、肩、肩、左肩周炎构建第二元路径,基于根痛平颗粒、肩颈疼痛、肩颈疼痛、左肩周炎构建第三元路径。

[0087] 在训练模型时,首先需要确定用于训练模型的滑动窗口尺寸,根据所确定的滑动窗口尺寸获取训练样本对,每组训练样本对包括输入样本和输出样本。根据训练样本对训练模型,从而得到模型的隐含层的参数。当模型训练好以后,并不会用这个训练好的模型处理新的任务,真正需要的是这个模型通过训练数据所学得的参数,以便获得文本的向量表示。利用训练模型对上述第一元路径、第二元路径和第三元路径进行训练,在构建并训练好词向量模型后,可以利用已训练好的词向量模型得到药品通用名向量(即药品向量)和诊断向量,该过程为离线操作。

[0088] 以图4(b)为例,输出为根痛平颗粒和左肩周炎两个的向量表示,分别是为 1×256 维度的向量。图谱中有多少个药品通用名、诊断的实体,输出就有多少个 1×256 的向量。所得药品通用名向量和诊断向量,可以存储在MySQL中,或者其他任意数据库中。可以预先设置药品通用名向量表和诊断向量表,药品通用名向量表用以存储药品通用名向量,诊断向量表用以存储诊断向量。两者不需要存储匹配关系,因为该匹配关系可以实时计算,计算方式就是两者的余弦相似度。余弦相似度是一个 $[-1, 1]$ 的数值,越趋近于1,表示两者存在治疗关系。

[0089] 上述实施例所提供的方法,将基于知识图谱的推荐算法引入到处方适应症审查场景中,以进一步挖掘药品通用名与诊断之间的关系,能够解决诊断描述多样化的问题。

[0090] 参见图5,示出了根据本发明实施例的另一种可选的数据审核方法流程示意图,包

括如下步骤：

[0091] S501：调用诊断属性提取模型，从诊断信息中提取第四属性、第五属性和第六属性；

[0092] S502：确定样本诊断信息库中与所述诊断信息的第四属性相似的第一样本诊断信息集合、与第五属性相似的第二样本诊断信息集合、以及与第六属性相似的第三样本诊断信息集合；

[0093] S503：对所述第一样本诊断信息集合、所述第二样本诊断信息集合和所述第三样本诊断信息集合做交集；

[0094] S504：响应于交集集合中的样本诊断信息数量仅为一个，得到与所述诊断信息相似的目标样本诊断信息；

[0095] S505：响应于交集集合中的样本诊断信息数量为多个，分别计算所述诊断信息和交集集合中每个样本诊断信息的相似度，将相似度最高的样本诊断信息，作为与所述诊断信息相似的目标样本诊断信息。

[0096] 上述实施方式中，对于步骤S501~S505，药品通用名通常为一个词，且根据前述描述，其为全世界通用词，通常不变，因而可以直接根据药品通用名从药品通用名向量表中查询到对应的药品通用名向量。但诊断信息通常涵盖信息较多，且考虑描述的多样性，即使针对同一症状的描述也会存有差异，因而需要采用一定的手段获取。

[0097] 实际操作中，可以计算本次诊断信息与样本诊断信息库中每个样本信息的文本相似度。也可以从两者记载的信息中提取关键词，以计算关键词之间的相似度，提升计算效率。将文本相似度最高的一个样本诊断信息，作为与本次诊断信息相似的目标样本诊断信息。

[0098] 根据前述描述，本方案主要考虑诊断信息的病理过程属性、发病部位属性和子症状属性，因而为进一步提升计算效率，还可以再次调用诊断属性提取模型，从本次诊断信息中提取这三个属性。同样考虑诊断信息描述的多样性，可能同样功效的诊断信息这三个属性会存在不同的情况，因而优选从这三个属性角度分别计算诊断信息属性相似度，例如疼痛和疼、肩膀和肩，得到三个集合：病理过程属性-第一样本诊断信息集合、发病部位属性-第二样本诊断信息集合、子症状属性-第三样本诊断信息集合。

[0099] 这三个集合的交集中，若仅存在一个样本诊断信息，则直接将其确定为与本次样本信息相似的样本诊断信息，之后从诊断向量表中查询与该样本诊断信息对应的诊断向量即可。但若存在两个及以上样本诊断信息，则采用文本相似度计算方式或关键词相似度计算方式，计算本次诊断信息与交集中每个样本诊断信息的相似度，最终筛选出相似度最高的一个样本诊断信息。

[0100] 上述实施例所提供的方法，通过属性相似度计算、文本相似度计算、关键词相似度计算方式，确定与本次诊断信息相似的样本诊断信息，直接查询该样本诊断信息的诊断向量，以进一步解决诊断信息描述多样性的问题，提升诊断向量查询的高效性和准确性。

[0101] 参见图6，示出了本发明实施例提供的一种数据审核装置600的主要模块示意图，包括：

[0102] 提取模块601，用于获取待审核的处方数据，从所述处方数据中提取药品信息和诊断信息，确定样本诊断信息库中与所述诊断信息相似的样本诊断信息；

[0103] 查询模块602,用于查询与所述药品信息对应的药品向量,以及查询与所述样本诊断信息对应的诊断向量;其中,药品向量和诊断向量通过元路径训练得到,元路径表示知识图谱中从药品信息到诊断信息之间的路径;

[0104] 计算模块603,用于计算所述药品向量和所述诊断向量的相似度,响应于所述相似度大于或等于预设相似度阈值,确定所述处方数据通过审核。

[0105] 本发明所述的数据审核装置还包括训练模块,用于:

[0106] 接收输入的样本药品诊断配对关系;其中,样本药品诊断配对关系包括匹配的样本药品信息和样本诊断信息;

[0107] 查询与所述样本药品信息对应的药品说明书信息,从所述药品说明书信息中提取药品特征信息,以及从所述样本诊断信息中提取诊断特征信息;

[0108] 基于所述药品特征信息和所述诊断特征信息构建知识图谱,使用所述知识图谱对机器学习模型进行训练,得到词向量模型,以基于所述词向量模型得到药品向量和诊断向量。

[0109] 本发明所述的数据审核装置中,所述训练模块用于:调用药品属性提取模型,从所述药品说明书信息中提取药品通用名、第一属性、第二属性和第三属性,以构建药品特征信息;

[0110] 所述训练模块用于:调用诊断属性提取模型,从所述样本诊断信息中提取诊断名称、第四属性、第五属性和第六属性,查询与所述第四属性对应的第七属性,以构建诊断特征信息。

[0111] 本发明所述的数据审核装置中,所述训练模块用于:

[0112] 基于药品通用名、第一属性、第七属性、第四属性、诊断名称,构建第一元路径;以及

[0113] 基于药品通用名、第二属性、第五属性、诊断名称,构建第二元路径;以及

[0114] 基于药品通用名、第三属性、第六属性、诊断名称,构建第三元路径;

[0115] 基于所述第一元路径、所述第二元路径和所述第三元路径构建知识图谱。

[0116] 本发明所述的数据审核装置中,所述提取模块601用于:

[0117] 分别计算所述诊断信息和样本诊断信息库中每个样本诊断信息的相似度,将相似度最高的样本诊断信息,作为与所述诊断信息相似的目标样本诊断信息。

[0118] 本发明所述的数据审核装置中,所述提取模块601,用于:

[0119] 调用诊断属性提取模型,从所述诊断信息中提取第四属性、第五属性和第六属性;

[0120] 确定所述样本诊断信息库中与所述诊断信息的第四属性相似的第一样本诊断信息集合、与第五属性相似的第二样本诊断信息集合、以及与第六属性相似的第三样本诊断信息集合;

[0121] 对所述第一样本诊断信息集合、所述第二样本诊断信息集合、所述第三样本诊断信息集合做交集,响应于交集集合中的样本诊断信息数量仅为一个,得到与所述诊断信息相似的目标样本诊断信息;或

[0122] 响应于交集集合中的样本诊断信息数量为多个,分别计算所述诊断信息和交集集合中每个样本诊断信息的相似度,将相似度最高的样本诊断信息,作为与所述诊断信息相似的目标样本诊断信息。

[0123] 另外,在本发明实施例中所述装置的具体实施内容,在上面所述方法中已经详细说明了,故在此重复内容不再说明。

[0124] 图7示出了可以应用本发明实施例的示例性系统架构700,包括终端设备701、702、703,网络704和服务器705(仅仅是示例)。

[0125] 终端设备701、702、703可以是具有显示屏并且支持网页浏览的各种电子设备,安装有各种通讯客户端应用,用户可以使用终端设备701、702、703通过网络704与服务器705交互,以接收或发送消息等。

[0126] 网络704用以在终端设备701、702、703和服务器705之间提供通信链路的介质。网络704可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0127] 服务器705可以是提供各种服务的服务器,需要说明的是,本发明实施例所提供的方法一般由服务器705执行,相应地,装置一般设置于服务器705中。

[0128] 应该理解,图7中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器。

[0129] 下面参考图8,其示出了适于用来实现本发明实施例的终端设备的计算机系统800的结构示意图。图8示出的终端设备仅仅是一个示例,不应对本发明实施例的功能和使用范围带来任何限制。

[0130] 如图8所示,计算机系统800包括中央处理单元(CPU)801,其可以根据存储在只读存储器(ROM)802中的程序或者从存储部分808加载到随机访问存储器(RAM)803中的程序而执行各种适当的动作和处理。在RAM 803中,还存储有系统800操作所需的各种程序和数据。CPU 801、ROM 802以及RAM 803通过总线804彼此相连。输入/输出(I/O)接口805也连接至总线804。

[0131] 以下部件连接至I/O接口805:包括键盘、鼠标等的输入部分806;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分807;包括硬盘等的存储部分808;以及包括诸如LAN卡、调制解调器等网络接口卡的通信部分809。通信部分809经由诸如因特网的网络执行通信处理。驱动器810也根据需要连接至I/O接口805。可拆卸介质811,诸如磁盘、光盘、磁光盘、半导体存储器等等,根据需要安装在驱动器810上,以便于从其上读出的计算机程序根据需要被安装入存储部分808。

[0132] 特别地,根据本发明公开的实施例,上文参考流程图描述的过程可以被实现为计算机软件程序。例如,本发明公开的实施例包括一种计算机程序产品,其包括承载在计算机可读介质上的计算机程序,该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中,该计算机程序可以通过通信部分809从网络上被下载和安装,和/或从可拆卸介质811被安装。在该计算机程序被中央处理单元(CPU)801执行时,执行本发明的系统中限定的上述功能。

[0133] 需要说明的是,本发明所示的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是一——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于:具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-ROM)、光存储器件、磁存储器件、

或者上述的任意合适的组合。在本发明中,计算机可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本发明中,计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质,该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于:无线、电线、光缆、RF等等,或者上述的任意合适的组合。

[0134] 附图中的流程图和框图,图示了按照本发明各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,上述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图或流程图中的每个方框、以及框图或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0135] 描述于本发明实施例中涉及到的模块可以通过软件的方式实现,也可以通过硬件的方式来实现。所描述的模块也可以设置在处理器中,例如,可以描述为:一种处理器包括提取模块、查询模块、计算模块。其中,这些模块的名称在某种情况下并不构成对该模块本身的限定,例如,查询模块还可以被描述为“向量查询模块”。

[0136] 作为另一方面,本发明还提供了一种计算机可读介质,该计算机可读介质可以是上述实施例中描述的设备中所包含的;也可以是单独存在,而未装配入该设备中。上述计算机可读介质承载有一个或者多个程序,当上述一个或者多个程序被一个该设备执行时,使得该设备执行上述任一所述的数据审核方法。

[0137] 本发明的计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现本发明实施例中的数据审核方法。

[0138] 上述具体实施方式,并不构成对本发明保护范围的限制。本领域技术人员应该明白的是,取决于设计要求和因素,可以发生各种各样的修改、组合、子组合和替代。任何在本发明的精神和原则之内所作的修改、等同替换和改进等,均应包含在本发明保护范围之内。

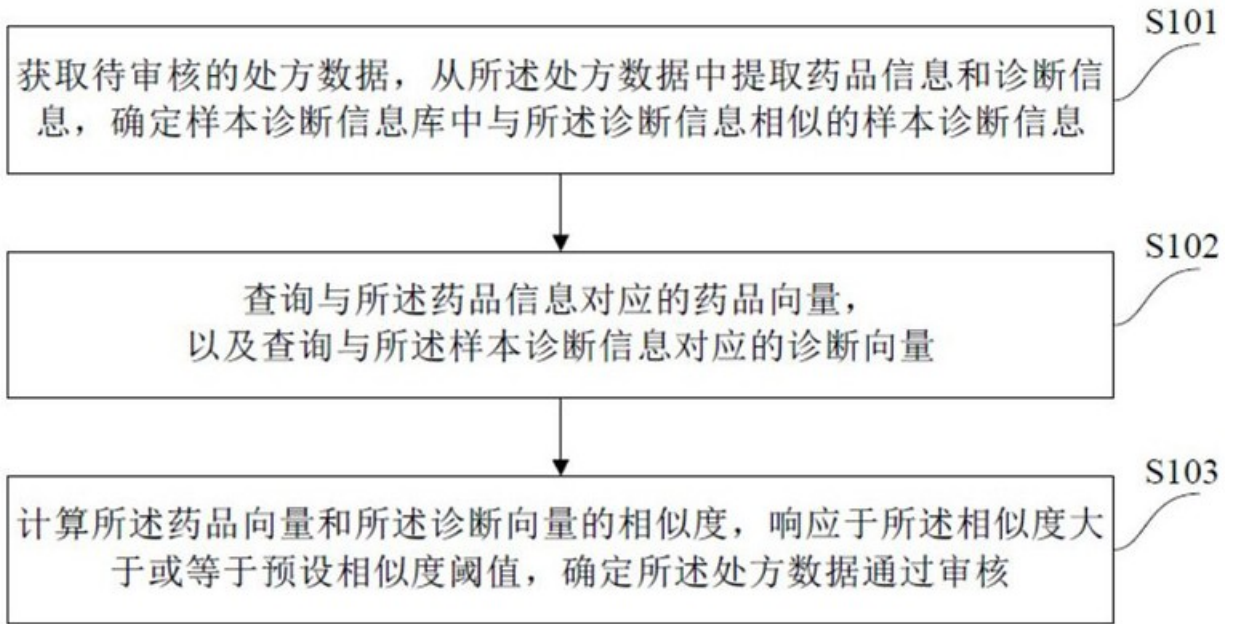


图1

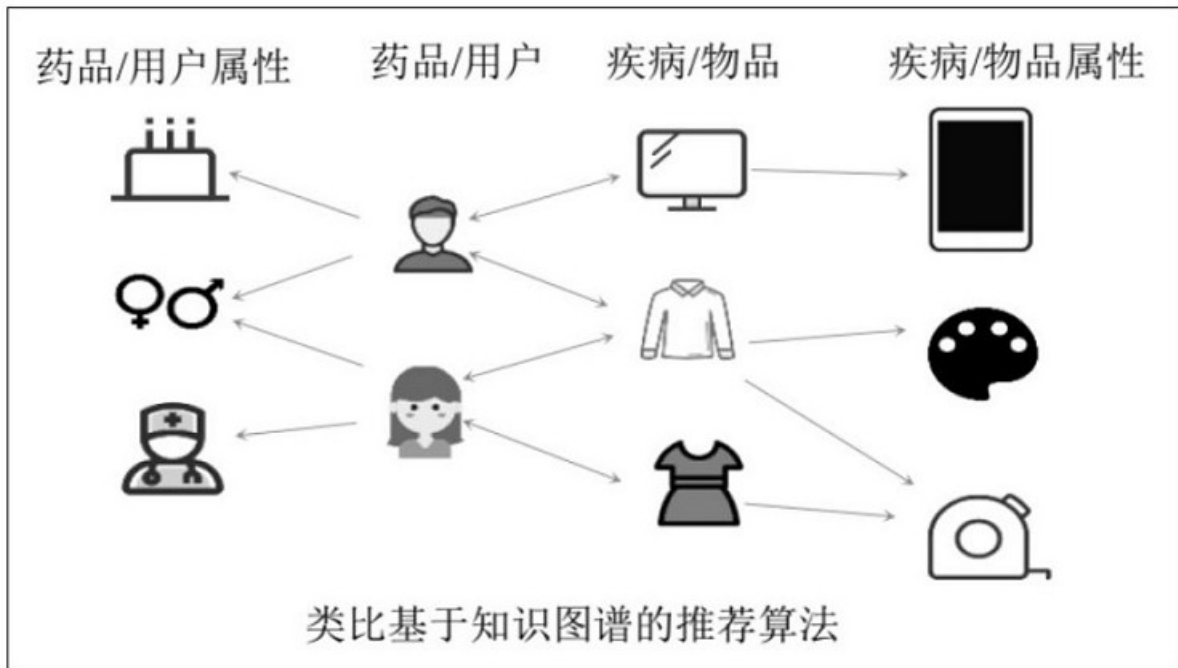


图2

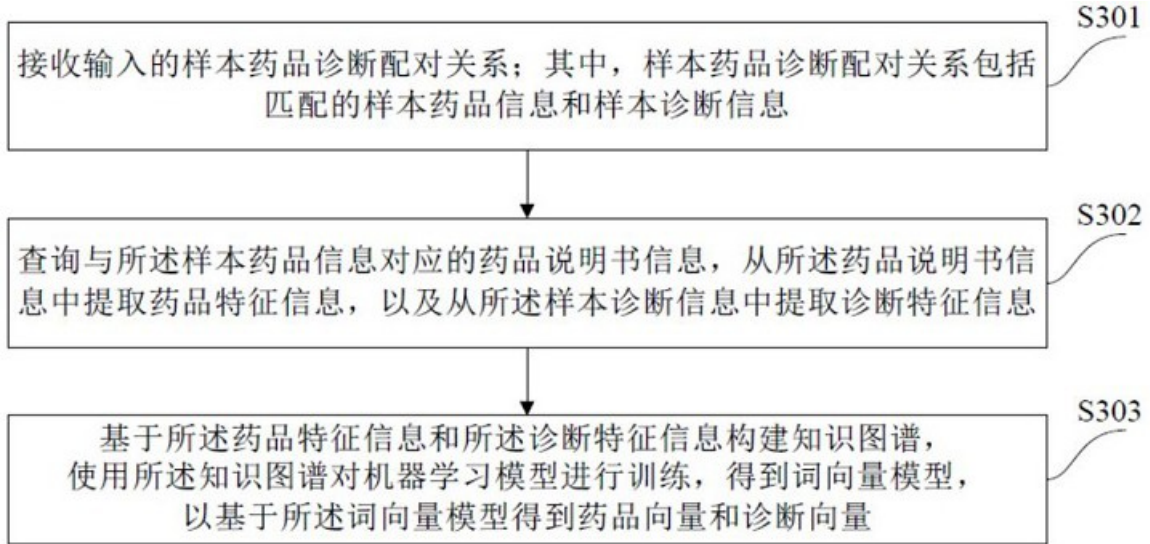


图3



图4(a)

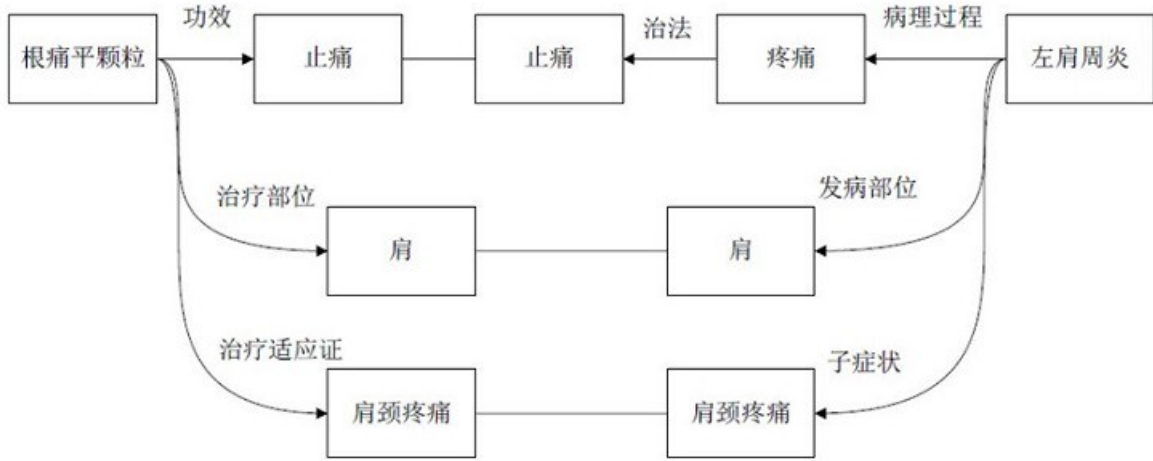


图4(b)

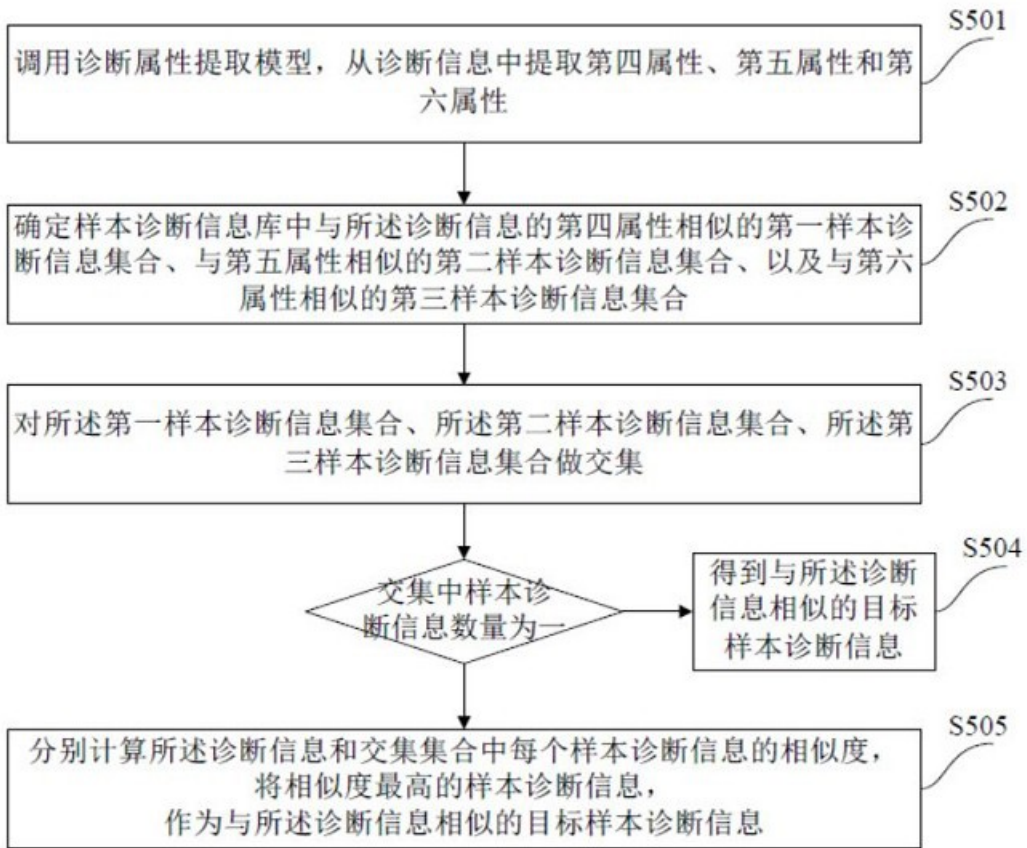


图5

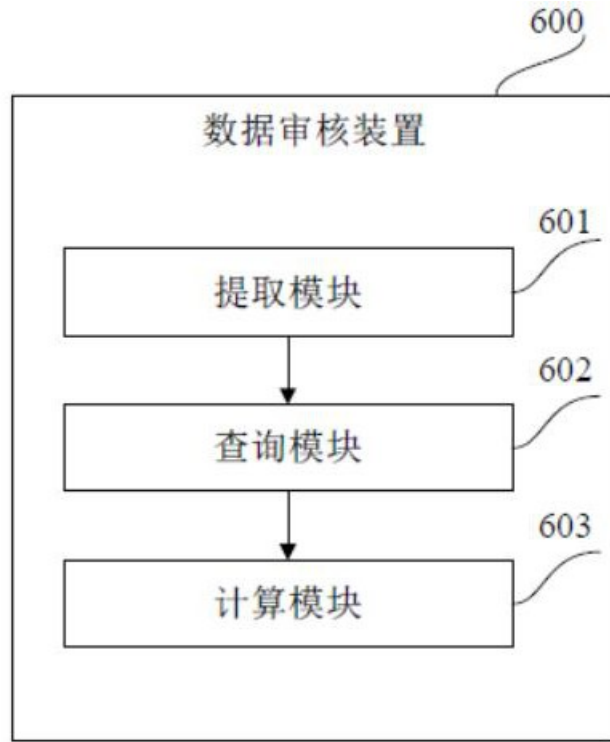


图6

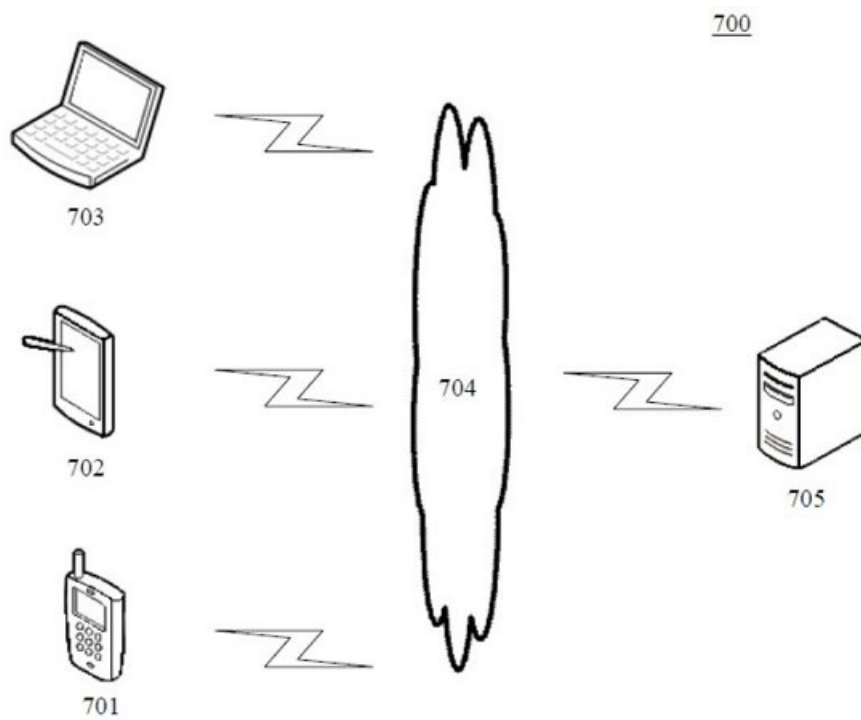


图7

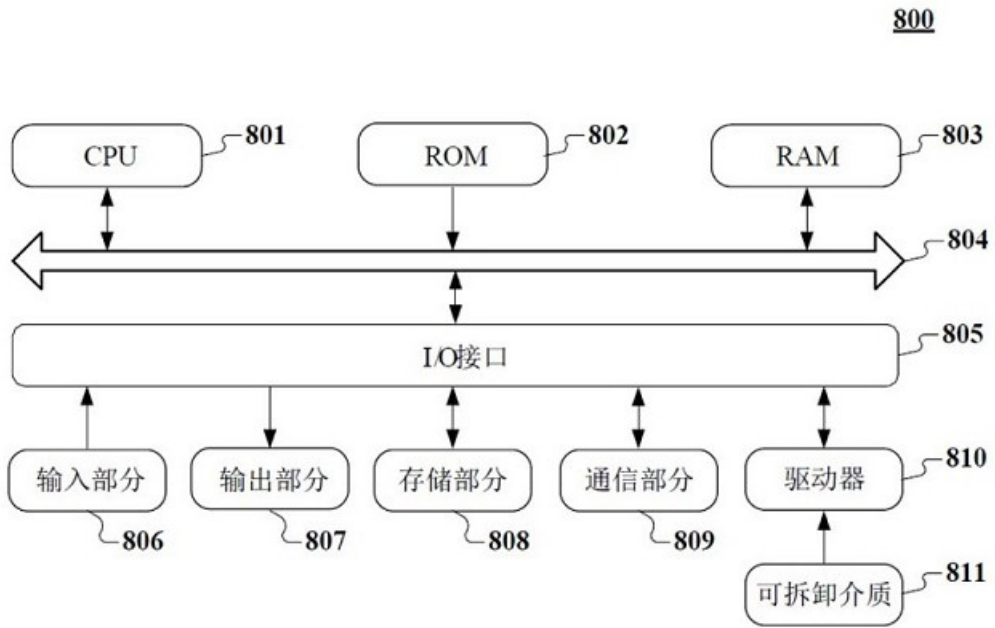


图8