



(12) 发明专利申请

(10) 申请公布号 CN 112599214 A

(43) 申请公布日 2021.04.02

(21) 申请号 202110237856.0

(22) 申请日 2021.03.04

(71) 申请人 中译语通科技股份有限公司
地址 100131 北京市石景山区石景山路20号中铁建设大厦16层

(72) 发明人 李勃兴 程国良

(74) 专利代理机构 北京兴智翔达知识产权代理有限公司 11768
代理人 郭卫芹

(51) Int. Cl.
G16H 10/60 (2018.01)
G06F 40/295 (2020.01)

权利要求书2页 说明书5页 附图1页

(54) 发明名称

从电子病历中提取出ICD编码的方法及系统

(57) 摘要

本发明公开了一种从电子病历中提取出ICD编码的方法及系统,该从电子病历中提取出ICD编码的方法包括:基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取,其中,所述医疗实体包括疾病诊断结果数据、检查数据、手术数据中的一个或多个;基于最小编辑距离算法确定出所述医疗实体所对应的ICD编码。本发明的从电子病历中提取出ICD编码的方法及系统能够快速提取出用于出院结算的ICD编码,帮助病人快速完成出院结算。



1. 一种从电子病历中提取出ICD编码的方法,其特征在于,包括:

基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取,其中,所述医疗实体包括疾病诊断结果数据、检查数据、手术数据中的一个或多个;以及基于最小编辑距离算法确定出所述医疗实体所对应的ICD编码。

2. 如权利要求1所述的从电子病历中提取出ICD编码的方法,其特征在于,所述基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取包括:

将所述电子病历中的病历小结的诊断结果中的逗号或分号后的数据作为医疗实体进行抽取;

基于深度学习序列标注的算法从所述电子病历的其余数据中抽取医疗实体。

3. 如权利要求1所述的从电子病历中提取出ICD编码的方法,其特征在于,所述基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取包括:

若基于所述深度学习序列标注的算法得到的所述电子病历中的医疗实体前存在否定词,则对该医疗实体不作抽取。

4. 如权利要求1所述的从电子病历中提取出ICD编码的方法,其特征在于,所述基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取包括:

在通过深度学习序列标注的算法对医疗实体进行抽取时,在标注数据时对所述电子病历的病史段出现的医疗实体作区别标注,对区别标注的医疗实体不作抽取。

5. 如权利要求1所述的从电子病历中提取出ICD编码的方法,其特征在于,所述基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取包括:

若基于所述深度学习序列标注的算法得到的所述电子病历中的医疗实体为病史段的数据,则对该病史段的医疗实体不作抽取。

6. 如权利要求1所述的从电子病历中提取出ICD编码的方法,其特征在于,所述基于最小编辑距离算法确定出所述医疗实体所对应的ICD编码包括:

对抽取的每个医疗实体,基于最小编辑距离算法分别求取所述每个医疗实体与ICD编码之间的相似度,若某个医疗实体与某个ICD编码之间的相似度达到预设阈值,则将所述某个医疗实体的数值归一化为所述某个ICD编码。

7. 如权利要求1所述的从电子病历中提取出ICD编码的方法,其特征在于,所述从电子病历中提取出ICD编码的方法还包括:

在对电子病历中的医疗实体进行抽取之后,基于疾病诊断结果数据与检查数据以及手术数据之间的对应关系对抽取的医疗实体进行校对,剔除错误数据;和/或

在基于最小编辑距离算法确定出所述医疗实体所对应的ICD编码之后,基于所述医疗实体的原始数据与所述ICD编码的对应关系,对所述ICD编码进行校对。

8. 一种从电子病历中提取出ICD编码的系统,其特征在于,包括:

实体抽取模块,用于基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取,其中,所述医疗实体包括疾病诊断结果数据、检查数据、手术数据中的多个;以及

ICD编码确定模块,与所述实体抽取模块相耦合,用于基于最小编辑距离算法确定出所述医疗实体所对应的ICD编码。

9. 一种电子装置,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算

机程序,其特征在于,所述处理器执行所述程序时实现如权利要求1至7任一项所述的从电子病历中提取出ICD编码的方法的步骤。

10.一种非暂态计算机可读存储介质,其上存储有计算机程序,其特征在于,该计算机程序被处理器执行时实现如权利要求1至7任一项所述的从电子病历中提取出ICD编码的方法的步骤。

从电子病历中提取出ICD编码的方法及系统

技术领域

[0001] 本发明是关于数据处理技术领域,特别是关于一种从电子病历中提取出ICD编码的方法及系统。

背景技术

[0002] ICD-9为手术用编码,ICD-10为疾病用编码。当前,国家为规范医疗行业电子病历,要求病历首页以ICD9和ICD10编码的方式反应病人在治疗过程中做过哪些检查,哪些手术并记录诊断结果。同时医保局将通过首页的编码记录进行报销。所以准确地提取病历中的诊断数据,检查数据以及手术数据(以下简称医疗实体)并对应到正确的ICD编码就变得尤为重要。

[0003] 目前大部医疗机构以人工的方式手工填写编码,工作人员需从几十页甚至上百页的电子病历中查找医疗实体,由于医生编写电子病历时的医疗实体并不完全符合编码对应的描述,可能会有错误,简写等情况,工作人员需要较高的知识储备才可进行一一对应。同时每天有大量病人出院结算,人工从电子病历中获取ICD编码的方式存在高技术门槛,工作量大以及大量重复工作等缺点。

[0004] 公开于该背景技术部分的信息仅仅旨在增加对本发明的总体背景的理解,而不应当被视为承认或以任何形式暗示该信息构成已为本领域一般技术人员所公知的现有技术。

发明内容

[0005] 本发明的目的在于提供一种从电子病历中提取出ICD编码的方法及系统,其能够快速提取出用于出院结算的ICD编码。

[0006] 为实现上述目的,本发明提供了一种从电子病历中提取出ICD编码的方法,其包括:基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取,其中,所述医疗实体包括疾病诊断结果数据、检查数据、手术数据中的一个或多个;基于最小编辑距离算法确定出所述医疗实体所对应的ICD编码。

[0007] 在本发明的一实施方式中,所述基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取包括:将所述电子病历中的病历小结的诊断结果中的逗号或分号后的数据作为医疗实体进行抽取;基于深度学习序列标注的算法从所述电子病历的其余数据中抽取医疗实体。

[0008] 在本发明的一实施方式中,所述基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取包括:若基于所述深度学习序列标注的算法得到的所述电子病历中的医疗实体前存在否定词,则对该医疗实体不作抽取。

[0009] 在本发明的一实施方式中,所述基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取包括:在通过深度学习序列标注的算法对医疗实体进行抽取时,在标注数据时对所述电子病历的病史段出现的医疗实体作区别标注,对区别标注的医疗实体不作抽取。

[0010] 在本发明的一实施方式中,所述基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取包括:若基于所述深度学习序列标注的算法得到的所述电子病历中的医疗实体为病史段的数据,则对该病史段的医疗实体不作抽取。

[0011] 在本发明的一实施方式中,所述基于最小编辑距离算法确定出所述医疗实体所对应的ICD编码包括:对抽取的每个医疗实体,基于最小编辑距离算法分别求取所述每个医疗实体与ICD编码之间的相似度,若某个医疗实体与某个ICD编码之间的相似度达到预设阈值,则将所述某个医疗实体的数值归一化为所述某个ICD编码。

[0012] 在本发明的一实施方式中,所述从电子病历中提取出ICD编码的方法还包括:在对电子病历中的医疗实体进行抽取之后,基于疾病诊断结果数据与检查数据以及手术数据之间的对应关系对抽取的医疗实体进行校对,剔除错误数据;和/或在基于最小编辑距离算法确定出所述医疗实体所对应的ICD编码之后,基于所述医疗实体的原始数据与所述ICD编码的对应关系,对所述ICD编码进行校对。

[0013] 基于同样的发明构思,本发明还提供了一种从电子病历中提取出ICD编码的系统,其包括:实体抽取模块以及ICD编码确定模块。实体抽取模块用于基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取,其中,所述医疗实体包括疾病诊断结果数据、检查数据、手术数据中的多个。ICD编码确定模块与所述实体抽取模块相耦合,用于基于最小编辑距离算法确定出所述医疗实体所对应的ICD编码。

[0014] 基于同样的发明构思,本发明还提供了一种电子装置,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现如上述任一项所述的从电子病历中提取出ICD编码的方法的步骤。

[0015] 基于同样的发明构思,本发明还提供了一种非暂态计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现如上述任一项所述的从电子病历中提取出ICD编码的方法的步骤。

[0016] 与现有技术相比,根据本发明的从电子病历中提取出ICD编码的方法及系统,针对非结构化的电子病历信息,基于序列标注算法以及预设规则对医疗实体进行自动化抽取,并且基于最小编辑距离算法确定医疗实体与ICD编码的相似度,从而在非结构化的数据快速准确抽取ICD编码信息并进行归一化处理,大幅降低了人工的工作量,可以帮助病人快速完成出院结算。

附图说明

[0017] 图1是根据本发明一实施方式的从电子病历中提取出ICD编码的方法;
图2是根据本发明一实施方式的从电子病历中提取出ICD编码的系统。

具体实施方式

[0018] 下面结合附图,对本发明的具体实施方式进行详细描述,但应当理解本发明的保护范围并不受具体实施方式的限制。

[0019] 除非另有其它明确表示,否则在整个说明书和权利要求书中,术语“包括”或其变换如“包含”或“包括有”等等将被理解为包括所陈述的元件或组成部分,而并未排除其它元件或其它组成部分。

[0020] 为了能够快速地从电子病历中提取出ICD编码,方便病人出院结算,一实施方式中提供了一种从电子病历中提取出ICD编码的方法,如图1所示,该方法包括:步骤S1~步骤S2。

[0021] 在步骤S1中,基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取,其中,所述医疗实体包括疾病诊断结果数据、检查数据、手术数据中的多个。

[0022] 具体而言,预设规则可以根据电子病历上的规律进行设定。如有些病例小结中‘诊断结果’后会以逗号或分号切分出患者做过哪些诊断。该种情况下,所述基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取包括:将所述电子病历中的病历小结的诊断结果中的逗号或分号后的数据作为医疗实体进行抽取;基于深度学习序列标注的算法从所述电子病历的其余数据中抽取医疗实体。由此,可以利用该规则进一步提高实体抽取的效率。

[0023] 再如,利用深度学习序列标注的算法识别出来的某些手术或检查在电子病历中的真实意思可能是患者拒绝做该手术或检查,对于该种情况,所述基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取包括:若基于所述深度学习序列标注的算法得到的所述电子病历中的医疗实体前存在否定词,则对该医疗实体不作抽取。

[0024] 再如,电子病历中的医疗实体可能是患者的病史数据,而不是本次检查结果,因而不能用于出院结算。对于该种情况,有两种处理方式可供选择。在第一种处理方式中,所述基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取包括:若基于所述深度学习序列标注的算法得到的所述电子病历中的医疗实体为病史段的数据,则对该病史段的医疗实体不作抽取。在第二种处理方式中,所述基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取包括:在通过深度学习序列标注的算法对医疗实体进行抽取时,在标注数据时在所述电子病历的病史段出现的医疗实体作区别标注,对区别标注的医疗实体不作抽取。

[0025] 在步骤S2中,基于最小编辑距离算法确定出所述医疗实体所对应的ICD编码。

[0026] 具体而言,考虑到电子病例中医生的编写方式可能会与ICD编码中的不符,因此在本步骤对抽取的医疗实体进行归一化处理。对抽取的每个医疗实体,基于最小编辑距离算法分别求取所述每个医疗实体与ICD编码之间的相似度,若某个医疗实体与某个ICD编码之间的相似度达到预设阈值,则将所述某个医疗实体的数值归一化为所述某个ICD编码。其中,编辑距离计算是指两个词组之间,由一个转成另一个所需要的最少编辑操作次数。允许的编辑操作包括:将一个字符替换成另一个字符,插入一个字符或者删除一个字符,通过次数计算两个词组的相似度。当相似度达到预设阈值时认为抽取的医疗实体为某个ICD-9编码或ICD-10编码。

[0027] 优选地,为了提高ICD提取结果的准确性,在一实施方式中,所述从电子病历中提取出ICD编码的方法还包括:在对电子病历中的医疗实体进行抽取之后,基于疾病诊断结果数据与检查数据以及手术数据之间的对应关系对抽取的医疗实体进行校对,剔除错误数据。具体而言,事先收集各项疾病诊断结果数据与检查数据以及手术数据之间的对应关系,基于该对应关系校对抽取出的医疗实体中,诊断结果是否跟检查以及手术对应。

[0028] 优选地,为了提高ICD提取结果的准确性,在一实施方式中,所述从电子病历中提取出ICD编码的方法还包括:在基于最小编辑距离算法确定出所述医疗实体所对应的

ICD编码之后,基于所述医疗实体的原始数据与所述ICD编码的对应关系,对所述ICD编码进行校对。

[0029] 基于同样的发明构思,一实施方式中还提供了一种从电子病历中提取出ICD编码的系统,如图2所示,其包括:实体抽取模块10以及ICD编码确定模块11。

[0030] 实体抽取模块10用于基于深度学习序列标注的算法以及预设规则对电子病历中的医疗实体进行抽取,其中,所述医疗实体包括疾病诊断结果数据、检查数据、手术数据中的多个。

[0031] 具体而言,预设规则可以根据电子病历上的规律进行设定。如有些病例小结中‘诊断结果’后会以逗号或分号切分出患者做过哪些诊断。该种情况下,所述抽取模块用于将所述电子病历中的病历小结的诊断结果中的逗号或分号后的数据作为医疗实体进行抽取;以及基于深度学习序列标注的算法从所述电子病历的其余数据中抽取医疗实体。由此,可以利用该规则进一步提高实体抽取的效率。

[0032] 再如,利用深度学习序列标注的算法识别出来的某些手术或检查在电子病历中的真实意思可能是患者拒绝做该手术或检查,对于该种情况,所述抽取模块还用于:若基于所述深度学习序列标注的算法得到的所述电子病历中的医疗实体前存在否定词,则对该医疗实体不作抽取。

[0033] 再如,电子病历中的医疗实体可能是患者的病史数据,而不是本次检查结果,因而不能用于出院结算。对于该种情况,有两种处理方式可供选择。在第一种处理方式中,所述抽取模块用于若基于所述深度学习序列标注的算法得到的所述电子病历中的医疗实体为病史段的数据,则对该病史段的医疗实体不作抽取。在第二种处理方式中,所述抽取模块用于在通过深度学习序列标注的算法对医疗实体进行抽取时,在标注数据时在所述电子病历的病史段出现的医疗实体作区别标注,对区别标注的医疗实体不作抽取。

[0034] ICD编码确定模块11与所述实体抽取模块10相耦合,用于基于最小编辑距离算法确定出所述医疗实体所对应的ICD编码。所述ICD编码确定模块11具体用于对抽取的每个医疗实体,基于最小编辑距离算法分别求取所述每个医疗实体与ICD编码之间的相似度,若某个医疗实体与某个ICD编码之间的相似度达到预设阈值,则将所述某个医疗实体的数值归一化为所述某个ICD编码。

[0035] 优选地,为了提高ICD提取结果的准确性,在一实施方式中,所述实体抽取模块10还用于:在对电子病历中的医疗实体进行抽取之后,基于疾病诊断结果数据与检查数据以及手术数据之间的对应关系对抽取的医疗实体进行校对,剔除错误数据。具体而言,事先收集各项疾病诊断结果数据与检查数据以及手术数据之间的对应关系,基于该对应关系校对抽取出的医疗实体中,诊断结果是否跟检查以及手术对应。

[0036] 优选地,为了提高ICD提取结果的准确性,在一实施方式中,所述ICD编码确定模块11还用于:在基于最小编辑距离算法确定出所述医疗实体所对应的ICD编码之后,基于所述医疗实体的原始数据与所述ICD编码的对应关系,对所述ICD编码进行校对。

[0037] 基于同样的发明构思,一实施方式中还提供了一种电子装置,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时实现如上述任一项所述的从电子病历中提取出ICD编码的方法的步骤。

[0038] 基于同样的发明构思,一实施方式中还提供了一种非暂态计算机可读存储介质,

其上存储有计算机程序,该计算机程序被处理器执行时实现如上述任一项所述的从电子病历中提取出ICD编码的方法的步骤。

[0039] 综上,根据本实施方式的从电子病历中提取出ICD编码的方法及系统,针对非结构化的电子病历信息,基于序列标注算法以及预设规则对医疗实体进行自动化抽取,并且基于最小编辑距离算法确定医疗实体与ICD编码的相似度,从而在非结构化的数据快速准确抽取ICD编码信息并进行归一化处理,大幅降低了人工的工作量,可以帮助病人快速完成出院结算。

[0040] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0041] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0042] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0043] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0044] 前述对本发明的具体示例性实施方案的描述是为了说明和例证的目的。这些描述并非想将本发明限定为所公开的精确形式,并且很显然,根据上述教导,可以进行很多改变和变化。对示例性实施例进行选择 and 描述的目的在于解释本发明的特定原理及其实际应用,从而使得本领域的技术人员能够实现并利用本发明的各种不同的示例性实施方案以及各种不同的选择和改变。本发明的范围意在由权利要求书及其等同形式所限定。

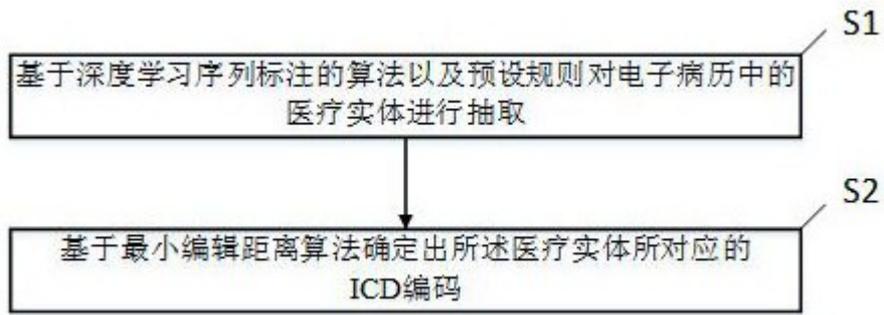


图1

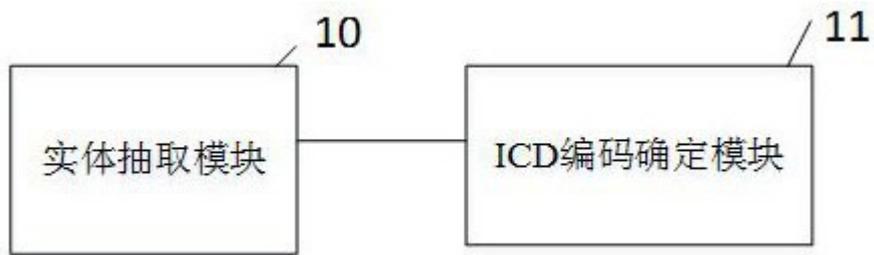


图2