



(12) 发明专利

(10) 授权公告号 CN 116168790 B

(45) 授权公告日 2023.07.14

(21) 申请号 202310450340.3

G16H 10/40 (2018.01)

(22) 申请日 2023.04.25

G16H 10/60 (2018.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 116168790 A

(56) 对比文件

US 2002172427 A1, 2002.11.21

(43) 申请公布日 2023.05.26

审查员 冯丽

(73) 专利权人 深圳爱递医药科技有限公司

地址 518000 广东省深圳市南山区粤海街道高新区社区高新南四道034号高新工业村W1-A座520

(72) 发明人 陈梓嘉 杨劲

(74) 专利代理机构 深圳政科创新专利代理事务

所(普通合伙) 44880

专利代理师 谢庚生

(51) Int. Cl.

G16H 10/20 (2018.01)

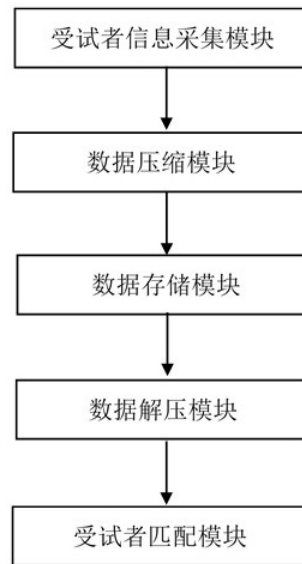
权利要求书2页 说明书7页 附图4页

(54) 发明名称

一种临床试验的大数据招募系统

(57) 摘要

本发明涉及数据处理领域,具体涉及一种临床试验的大数据招募系统,该系统包括:受试者信息采集模块、数据压缩模块、数据存储模块、数据解压模块、受试者匹配模块。根据受试者的医疗资料获取待压缩数据,获取待压缩数据中的前向像素与后向像素,根据前向像素与后向像素的关系以及前向像素是否在编码字典中存在,结合前向像素的游程长度对前向像素进行编码,获得压缩数据,对压缩数据进行存储,根据受试者的医疗资料筛选符合临床试验的条件的受试者参与临床试验。本发明结合游程对LZ编码进行改进,压缩效率更高,节省了受试者库中数据的存储空间。



1. 一种临床试验的大数据招募系统,其特征在于,所述系统包括:
受试者信息采集模块,根据受试者的医疗资料获取待压缩数据;
数据压缩模块,构建一个空的编码字典;将待压缩数据中第一个元素作为前向像素,第二个元素作为后向像素;
根据前向像素与后向像素进行压缩操作,包括:
S1:当前向像素的长度为1且前向像素与后向像素相同时,获取新的前向像素与新的后向像素;
S2:当前向像素的长度为1且前向像素与后向像素不同时,对前向像素进行第一判断操作;
S3:当前向像素的长度大于1且前向像素中仅包含了一种像素值时,对前向像素与后向像素进行第二判断操作,包括:
当后向像素与前向像素包含的像素值相同时,或后向像素与前向像素包含的像素值不相同但前向像素在编码字典中存在时,获取新的前向像素与新的后向像素;当后向像素与前向像素包含的像素值不相同且前向像素在编码字典中不存在时,根据编码字典与前向像素的长度获取第一编码,将前向像素中像素值对应的八位二进制数作为第二编码,根据第一编码与第二编码更新编码字典、前向像素与后向像素;
S4:当前向像素的长度大于1且前向像素中包含了多种像素值时,对前向像素进行第三判断操作;
重复根据新的前向像素与新的后向像素进行压缩操作,直到不存在新的前向像素时停止迭代;根据编码字典中所有编码获取压缩数据;
数据存储模块,将压缩数据存储在接受者库存储服务器;
数据解压模块,对压缩数据进行解压,得到受试者的医疗资料;
受试者匹配模块,根据所有受试者的医疗资料筛选符合临床试验的条件的受试者参与临床试验;
所述对前向像素进行第一判断操作,包括的步骤为:
判断前向像素在编码字典中是否存在:
当前向像素在编码字典中不存在时,将前向像素转换为八位二进制数并拼接在0之后,将所得的结果作为前向像素的编码,将前向像素以及前向像素的编码添加到编码字典中,将后向像素作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素;
当前向像素在编码字典中存在时,将前向像素与后向像素拼接在一起作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素;
所述对前向像素进行第三判断操作,包括的步骤为:
判断前向像素在编码字典中是否存在:
当前向像素在编码字典中不存在时,获取前向像素的长度N,获取前向像素中前N-1个元素在编码字典中所在行的行号,将行号对应的二进制数作为第一编码,获取前向像素中最后一个元素对应的八位二进制数作为第二编码,将第二编码拼接在第一编码之后作为前向像素的编码,将前向像素以及前向像素的编码添加到编码字典中,将后向像素作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素;

当前向像素在编码字典中存在时,将前向像素与后向像素拼接在一起作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素。

2. 根据权利要求1所述的一种临床试验的大数据招募系统,其特征在于,所述根据受试者的医疗资料获取待压缩数据,包括的步骤为:

获取受试者的医疗资料中的医疗影像,将医疗影像中所有像素点的像素值构成一个序列,作为待压缩数据。

3. 根据权利要求1所述的一种临床试验的大数据招募系统,其特征在于,所述获取新的前向像素与新的后向像素,包括的步骤为:

将前向像素与后向像素拼接在一起作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素。

4. 根据权利要求1所述的一种临床试验的大数据招募系统,其特征在于,所述根据编码字典与前向像素的长度获取第一编码,包括的步骤为:

获取编码字典中最后一行的行号,获取行号与前向像素的长度的加和,将所述加和对应的二进制数作为第一编码。

5. 根据权利要求1所述的一种临床试验的大数据招募系统,其特征在于,所述根据第一编码与第二编码更新编码字典、前向像素与后向像素,包括的步骤为:

将第二编码拼接在第一编码之后作为前向像素的编码,将前向像素以及前向像素的编码添加到编码字典中,将后向像素作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素。

6. 根据权利要求1所述的一种临床试验的大数据招募系统,其特征在于,所述根据编码字典中所有编码获取压缩数据,包括的步骤为:

获取编码字典中所有编码的长度中的最大值L,在编码字典中每个编码前补0使每个编码的长度达到L;将编码字典中所有编码按照顺序拼接在一起得到一个二进制序列,将所述二进制序列作为压缩数据。

7. 根据权利要求1所述的一种临床试验的大数据招募系统,其特征在于,所述对压缩数据进行解压,得到受试者的医疗资料,包括的步骤为:

构建一个空的编码字典;将压缩数据划分为多个二进制数,将每个二进制数作为一个编码,将所有编码按照顺序填入到编码字典的编码列;从编码字典中第一个编码开始,依次对每个编码进行解压,包括:

根据编码的长度将编码划分为第一编码以及第二编码,当第一编码对应的十进制数P为0时,将第二编码对应的十进制数作为编码的解压结果;当P大于0且小于编码在编码字典中所在行的行号M时,将编码字典中行号为P的编码对应的解压结果作为第一短语,将第二编码对应的十进制数作为第二短语,将第二短语拼接在第一短语之后,得到的结果作为编码的解压结果;当P大于M时,将 $P - (M - 1)$ 作为游程长度,将第二编码对应的十进制数作为第一像素值,根据游程长度以及第一像素值获取编码的解压结果;

将编码字典中每个编码的解压结果按照顺序拼接在一起,得到解压序列,根据解压序列获取受试者的医疗资料中的医疗影像。

一种临床试验的大数据招募系统

技术领域

[0001] 本发明涉及数据处理领域,具体涉及一种临床试验的大数据招募系统。

背景技术

[0002] 临床试验,指任何在人体(病人或健康志愿者)进行药物的系统性研究,以证实或揭示试验药物的作用、不良反应或试验药物的吸收、分布、代谢和排泄,目的是确定试验药物的疗效与安全性。

[0003] 目前临床试验招募受试者主要是由人工根据受试者的病例、检查报告等筛选匹配的受试者,或收集受试者的病例、检查报告等信息建立受试者库,通过大数据自动在受试者库中匹配适合的受试者。

[0004] 由于受试者库中的数据庞大,需要进行压缩存储,现有的压缩方法如LZ压缩方法可通过将重复出现的两个或两个以上的字符编码成一个码字来达到压缩的效果,但对于长游程的数据,LZ编码会将其编码成多个码字,此时难以达到最高的压缩效果。

发明内容

[0005] 为了解决上述问题,本发明提供一种临床试验的大数据招募系统,所述系统包括:

[0006] 受试者信息采集模块,根据受试者的医疗资料获取待压缩数据;

[0007] 数据压缩模块,构建一个空的编码字典;将待压缩数据中第一个元素作为前向像素,第二个元素作为后向像素;

[0008] 根据前向像素与后向像素进行压缩操作,包括:

[0009] S1:当前向像素的长度为1且前向像素与后向像素相同时,获取新的前向像素与新的后向像素;

[0010] S2:当前向像素的长度为1且前向像素与后向像素不同时,对前向像素进行第一判断操作;

[0011] S3:当前向像素的长度大于1且前向像素中仅包含了一种像素值时,对前向像素与后向像素进行第二判断操作,包括:

[0012] 当后向像素与前向像素包含的像素值相同时,或后向像素与前向像素包含的像素值不相同但前向像素在编码字典中存在时,获取新的前向像素与新的后向像素;当后向像素与前向像素包含的像素值不相同且前向像素在编码字典中不存在时,根据编码字典与前向像素的长度获取第一编码;将前向像素中像素值对应的八位二进制数作为第二编码,根据第一编码与第二编码更新编码字典、前向像素与后向像素;

[0013] S4:当前向像素的长度大于1且前向像素中包含了多种像素值时,对前向像素进行第三判断操作;

[0014] 重复根据新的前向像素与新的后向像素进行压缩操作,直到不存在新的前向像素时停止迭代;根据编码字典中所有编码获取压缩数据;

[0015] 数据存储模块,将压缩数据存储于受试者库存储服务器;

- [0016] 数据解压模块,对压缩数据进行解压,得到受试者的医疗资料;
- [0017] 受试者匹配模块,根据所有受试者的医疗资料筛选符合临床试验的条件的受试者参与临床试验。
- [0018] 优选的,所述根据受试者的医疗资料获取待压缩数据,包括的步骤为:
- [0019] 获取受试者的医疗资料中的医疗影像,将医疗影像中所有像素点的像素值构成一个序列,作为待压缩数据。
- [0020] 优选的,所述获取新的前向像素与新的后向像素,包括的步骤为:
- [0021] 将前向像素与后向像素拼接在一起作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素。
- [0022] 优选的,所述对前向像素进行第一判断操作,包括的步骤为:
- [0023] 判断前向像素在编码字典中是否存在:
- [0024] 当前向像素在编码字典中不存在时,将前向像素转换为八位二进制数并拼接在0之后,将所得的结果作为前向像素的编码,将前向像素以及前向像素的编码添加到编码字典中,将后向像素作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素;
- [0025] 当前向像素在编码字典中存在时,将前向像素与后向像素拼接在一起作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素。
- [0026] 优选的,所述根据编码字典与前向像素的长度获取第一编码,包括的步骤为:
- [0027] 获取编码字典中最后一行的行号,获取行号与前向像素的长度的加和,将所述加和对应的二进制数作为第一编码。
- [0028] 优选的,所述根据第一编码与第二编码更新编码字典、前向像素与后向像素,包括的步骤为:
- [0029] 将第二编码拼接在第一编码之后作为前向像素的编码,将前向像素以及前向像素的编码添加到编码字典中,将后向像素作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素。
- [0030] 优选的,所述对前向像素进行第三判断操作,包括的步骤为:
- [0031] 判断前向像素在编码字典中是否存在:
- [0032] 当前向像素在编码字典中不存在时,获取前向像素的长度N,获取前向像素中前N-1个元素在编码字典中所在行的行号,将行号对应的二进制数作为第一编码,获取前向像素中最后一个元素对应的八位二进制数作为第二编码,将第二编码拼接在第一编码之后作为前向像素的编码,将前向像素以及前向像素的编码添加到编码字典中,将后向像素作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素;
- [0033] 当前向像素在编码字典中存在时,将前向像素与后向像素拼接在一起作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素。
- [0034] 优选的,所述根据编码字典中所有编码获取压缩数据,包括的步骤为:
- [0035] 获取编码字典中所有编码的长度中的最大值L,在编码字典中每个编码前补0使每个编码的长度达到L;将编码字典中所有编码按照顺序拼接在一起得到一个二进制序列,将所述二进制序列作为压缩数据。
- [0036] 优选的,所述对压缩数据进行解压,得到受试者的医疗资料,包括的步骤为:

[0037] 构建一个空的编码字典;将压缩数据划分为多个二进制数,将每个二进制数作为一个编码,将所有编码按照顺序填入到编码字典的编码列;从编码字典中第一个编码开始,依次对每个编码进行解压,包括:

[0038] 根据编码的长度将编码划分为第一编码以及第二编码,当第一编码对应的十进制数P为0时,将第二编码对应的十进制数作为编码的解压结果;当P大于0且小于编码在编码字典中所在行的行号M时,将编码字典中行号为P的编码对应的解压结果作为第一短语,将第二编码对应的十进制数作为第二短语,将第二短语拼接在第一短语之后,得到的结果作为编码的解压结果;当P大于M时,将P-(M-1)作为游程长度,将第二编码对应的十进制数作为第一像素值,根据游程长度以及第一像素值获取编码的解压结果;

[0039] 将编码字典中每个编码的解压结果按照顺序拼接在一起,得到解压序列,根据解压序列获取受试者的医疗资料中的医疗影像。

[0040] 本发明实施例至少具有如下有益效果:现有的压缩方法如LZ压缩方法可通过将重复出现的两个或两个以上的字符编码成一个码字来达到压缩的效果,但对于长游程的数据,LZ编码会将其编码成多个码字,难以达到最高的压缩效果。本发明通过将获取待压缩数据中的前向像素,在LZ编码的基础上,对于仅包含一种像素值的前向像素,即长游程的像素值,利用游程长度与编码字典的中最后一行的行号之和获取第一编码,根据第一编码以及前向像素中包含的像素值获取长游程的数据的编码,从而实现将长游程的数据编码成一个码字,相较于LZ编码大大提升了压缩效率,节省了受试者库中数据的存储空间。

附图说明

[0041] 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和优点,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的介绍,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它附图。

[0042] 图1为本发明一个实施例提供的一种临床试验的大数据招募系统的系统框图;

[0043] 图2为医疗影像;

[0044] 图3为待压缩数据示意图;

[0045] 图4为LZ编码的编码字典;

[0046] 图5为本发明中压缩方法的编码字典;

[0047] 图6为LZ编码的压缩结果;

[0048] 图7为本发明中压缩方法的压缩结果。

具体实施方式

[0049] 为了更进一步阐述本发明为达成预定发明目的所采取的技术手段及功效,以下结合附图及较佳实施例,对依据本发明提出的一种临床试验的大数据招募系统,其具体实施方式、结构、特征及其功效,详细说明如下。在下述说明中,不同的“一个实施例”或“另一个实施例”指的不一定是同一实施例。此外,一或多个实施例中的特定特征、结构、或特点可由任何合适形式组合。

[0050] 除非另有定义,本文所使用的所有的技术和科学术语与属于本发明的技术领域的

技术人员通常理解的含义相同。

[0051] 下面结合附图具体的说明本发明所提供的一种临床试验的大数据招募系统的具体方案。

[0052] 请参阅图1,其示出了本发明一个实施例提供的一种临床试验的大数据招募系统,该系统包括以下模块:

[0053] 受试者信息采集模块用于获取待压缩数据。

[0054] 具体的,临床试验志愿者将医疗资料,如个人病例、检查报告、医疗影像等上传至临床试验招募系统。招募系统将志愿者作为受试者加入受试者库,并将受试者的医疗资料存储至受试者库存储服务器。由于受试者库数据庞大,因此需要对受试者的医疗资料进行压缩存储。本发明实施例针对受试者的医疗资料中的医疗影像进行压缩,因此获取受试者的医疗影像,根据医疗影像获取待压缩数据,并记录医疗影像的大小。本发明实施例中的一个医疗影像参见图2。

[0055] 需要说明的是,根据医疗影像获取待压缩数据的具体方法为:

[0056] 部分医疗影像如CT影像为灰度图像,此时则将所有像素点的灰度值构成一个一维的序列,作为待压缩数据;

[0057] 部分医疗影像如彩色多普勒超声显像为彩色图像,此时则将所有像素点的每个通道的像素值分别组成一个一维的序列,并将所有的一维的序列拼接成一个序列,作为待压缩数据。例如根据图2所示的医疗影像获得的待压缩数据参见图3。

[0058] 至此,获取待压缩数据。

[0059] 数据压缩模块用于对待压缩数据进行压缩。

[0060] 需要说明的是,现有的压缩方法如LZ算法对图像进行压缩时,通过将重复出现两个及以上的像素值编码成一个码字来实现图像的压缩。但由于医疗影像具有局部相似性,因此医疗影像中像素值的游程较长,即待压缩数据中像素值的游程较长。而利用LZ算法会将长游程的像素值编码成多个码字,此时压缩效率不高。本发明实施例结合医疗影像的长游程特征对LZ算法进行改进,使得待压缩数据的压缩效率更高。

[0061] 在本发明实施例中,首先构建一个空的编码字典,用于在对待压缩数据进行压缩时动态存放待压缩数据中的一个或多个像素值的压缩结果。编码字典第一列为段号,用来记录编码字典的行号;第二列为短语,用来记录被编码的内容;第三列为编码,用来记录编码的结果。

[0062] 将待压缩数据中第一个元素作为前向像素,将待压缩数据中第二个元素作为后向像素,根据前向像素与后向像素进行压缩操作,包括:

[0063] 1、当前向像素的长度为1且前向像素与后向像素相同时,将前向像素与后向像素拼接在一起作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素;

[0064] 2、当前向像素的长度为1且前向像素与后向像素不同时,判断前向像素在编码字典中短语列是否存在:

[0065] 当前向像素在编码字典中短语列不存在时,将前向像素转换为八位二进制数,并将该八位二进制数拼接在0之后,将所得的结果作为前向像素的编码,此时将前向像素以及前向像素的编码添加到编码字典中,其中将前向像素添加到编码字典新的一行的第二列

(即短语列),将前向像素的编码添加到编码字典新的一行的第三列(即编码列),此时将后向像素作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素;

[0066] 当前向像素在编码字典中短语列存在时,将前向像素与后向像素拼接在一起作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素;

[0067] 3、当前向像素的长度大于1且前向像素中仅包含了一种像素值时(本发明实施例中相同的像素值为同一种像素值),判断后向像素是否与前向像素包含的像素值相同;

[0068] 当后向像素与前向像素包含的像素值相同时,将前向像素与后向像素拼接在一起作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素;

[0069] 当后向像素与前向像素包含的像素值不相同且前向像素在编码字典中存在时,将前向像素与后向像素拼接在一起作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素;

[0070] 当后向像素与前向像素包含的像素值不相同且前向像素在编码字典中不存在时,获取编码字典中存在的最大段号S(即编码字典中最后一行的行号),以及前向像素的长度N,获取S+N对应的二进制数作为第一编码,获取前向像素中像素值对应的八位二进制数作为第二编码,将第二编码拼接在第一编码之后作为前向像素的编码,此时将前向像素以及前向像素的编码添加到编码字典中,其中将前向像素添加到编码字典新的一行的第二列(即短语列),将前向像素的编码添加到编码字典新的一行的第三列(即编码列),此时将后向像素作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素;

[0071] 4、当前向像素的长度大于1且前向像素中包含了多种像素值时,判断前向像素在编码字典中是否存在:

[0072] 当前向像素在编码字典中短语列不存在时,获取前向像素的长度N,获取前向像素中前N-1个元素组成的短语在编码字典中对应的段号(即前向像素中前N-1个元素组成的短语在编码字典中所在行的行号),将该段号对应的二进制数作为第一编码,获取前向像素中最后一个元素对应的八位二进制数作为第二编码,将第二编码拼接在第一编码之后作为前向像素的编码,此时将前向像素以及前向像素的编码添加到编码字典中,其中将前向像素添加到编码字典新的一行的第二列(即短语列),将前向像素的编码添加到编码字典新的一行的第三列(即编码列),此时将后向像素作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素;

[0073] 当前向像素在编码字典中短语列存在时,将前向像素与后向像素拼接在一起作为新的前向像素,将待压缩数据中后向像素的下一个元素作为新的后向像素;

[0074] 重复根据新的前向像素与新的后向像素进行压缩操作,直到不存在新的前向像素时停止迭代。

[0075] 获取编码字典中所有编码的长度中的最大值L,在编码字典中每个编码前补0使每个编码的长度达到L,如此使得编码字典中所有编码的长度一致。将编码字典中所有编码按照顺序拼接在一起得到一个一维的二进制序列,将该二进制序列作为压缩数据。

[0076] 至此,获取了压缩数据。

[0077] 需要说明的是,利用现有的LZ编码对{172,32,32,32,54,32,32,32,54,172,173}进行压缩得到的编码字典如图4所示,得到的压缩数据为{0001010110000000100000010001000000000110110011001000001001010110000010101101}。由于LZ编码对首次出现的像

素值编码成一个码字,对再次出现的像素值与该像素值之后的像素值组成的短语编码成一个码字,如此导致了将长游程的数据编码成多个码字,例如LZ编码将{172,32,32,32,54,32,32,32,54,172,173}中的首次出现的32编码成00000100000,将再次出现的32与第三个32组成的短语32,32编码成01000100000。

[0078] 而采用本发明实施例中的方法对{172,32,32,32,54,32,32,32,54,172,173}进行压缩得到的编码字典如图5所示,得到的压缩数据为{0001010110010000100000000001101100100011011000110101101},本发明实施例中结合游程的特点进行编码,将{172,32,32,32,54,32,32,32,54,172,173}中首次出现的32,32,32编码成10000100000,相较于LZ编码将首次出现的32,32,32编码成两个码字00000100000、01000100000,本发明的压缩效率更高。

[0079] 对{172,32,32,32,54,32,32,32,54,172,173}利用LZ编码得到的压缩数据的长度为77,利用本发明实施例中的方法得到的压缩数据的长度为55。由此可知,本发明实施例中的方法压缩效率相较LZ编码的压缩效率更高。

[0080] 利用对图3所示的待压缩数据利用LZ编码进行压缩得到的部分结果参见图6,对图3所示的待压缩数据利用本发明实施例中的方法进行压缩得到的部分结果参见图7,对图2中医疗影响利用不同压缩方法压缩的压缩率参见表1。

[0081] 表1压缩率对比表

压缩方法	原始数据量	压缩后数据量	压缩率
本发明算法	81152byte	26557byte	32.73%
LZ算法	81152byte	39845byte	49.10%

[0082] 数据存储模块用于存储压缩数据。

[0083] 具体的,招募系统将压缩数据存储至受试者库存储服务器,同时还需保存编码字典中编码的长度L以及医疗影像的大小。

[0084] 数据解压模块用于读取并解压压缩数据。

[0085] 需要说明的是,当需要对受试者库中的受试者进行临床试验筛选时,需要根据受试者的个人病例、检查报告进行大数据匹配,此时需要先对压缩数据进行解压,以获取受试者的医疗资料中的医疗影像。

[0086] 在本发明实施例中,对压缩数据进行解压的具体方法为:

[0087] 首先构建一个空的编码字典,用于在对压缩数据进行解压时动态存放压缩数据对应的像素值。编码字典第一列为段号,用来记录编码字典的行数;第二列为短语,用来记录被编码的内容;第三列为编码,用来记录编码的结果。

[0088] 将压缩数据划分为多个长度为L的二进制数,将每个二进制数作为一个编码,将所有编码按照顺序填入到编码字典的第三列(即编码列)。从编码字典中第一个编码开始,依次对每个编码进行解压,具体为:

[0089] 编码的长度为L,将编码的第1个至第L-8个元素作为第一编码,将编码的后8个元素作为第二编码。获取编码所在行的段号M,将第一编码转换为十进制数,并记为P,根据M和P对编码进行解压:

[0090] 当P=0时,将第二编码转换为十进制数,作为编码的解压结果,将编码的解压结果

填入到编码所在行的第二列(即短语列);

[0092] 当 $0 < P < M$ 时,获取编码字典中段号为P的行中短语列的内容,将该内容作为第一短语,将第二编码转换为十进制数,作为第二短语,将第二短语拼接在第一短语之后,得到的结果作为编码的解压结果,将编码的解压结果填入到编码所在行的第二列(即短语列);

[0093] 当 $P > M$ 时,将 $P - (M - 1)$ 作为游程长度,将第二编码转换为十进制数,作为第一像素值,记为B,将 $P - (M - 1)$ 个第一像素值B拼接在一起,作为编码的解压结果,将编码的解压结果填入到编码所在行的第二列(即短语列)。

[0094] 如此可获得完整的编码字典,将编码字典中第二列(即短语列)中的内容按照顺序拼接在一起,得到的结果即为医疗影像中所有像素点的像素值所构成的序列,将其记为解压序列,该解压序列即受试者信息采集模块中的待压缩数据。

[0095] 根据医疗影像的大小获取医疗影像中包含的像素点个数,如医疗影像的长为a,宽为b时,医疗影像中包含的像素点个数为 $a \times b$ 。当解压序列的长度与医疗影像中包含的像素点个数一致时,解压序列中每个元素对应医疗影像中每个像素点的像素值,此时根据解压序列以及医疗影像的大小还原医疗影像;当解压序列的长度大于医疗影像中包含的像素点个数时,解压序列中元素为医疗影像中每个像素点多个通道的像素值,将解压序列分割成多个长度为医疗影像中包含的像素点个数的一维序列,每个一维序列中每个元素为解压序列中每个像素点的每个通道的像素值,此时根据多个一维序列以及医疗影像的大小还原医疗影像。

[0096] 至此,完成了对压缩数据的解压,获取了受试者的医疗影像。

[0097] 受试者匹配模块用于筛选匹配临床试验的受试者。

[0098] 招募系统依据临床试验的适用范围,根据受试者库中所有受试者的医疗资料,如个人病例、检查报告以及医疗影像等信息对受试者进行自动筛选匹配,筛选出符合临床试验的条件的受试者参与临床试验。

[0099] 至此,完成了临床试验的受试者筛选。

[0100] 综上所述,本发明的系统包括受试者信息采集模块、数据压缩模块、数据存储模块、数据解压模块、受试者匹配模块,本发明通过将获取待压缩数据中的前向像素,在LZ编码的基础上,对于仅包含一种像素值的前向像素,即长游程的数据,利用游程长度与编码字典的中最后一行的行号之和获取第一编码,根据第一编码以及前向像素中包含的像素值获取长游程的数据的编码,从而实现将长游程的数据编码成一个码字,相较于LZ编码大大提升了压缩效率,节省了受试者库中数据的存储空间。

[0101] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

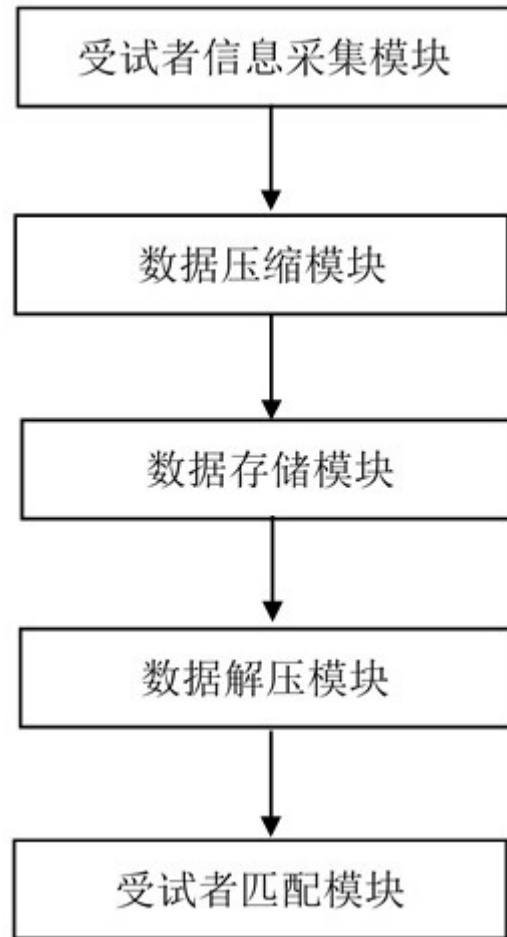


图1



图2

11,10,10,10,9,9,9,9,9,9,9,9,9,9,9,9,9,8,8,8,8,8,8,9,9,10,11,11,9,8,8,10,11
,10,10,13,16,15,15,17,36,120,148,123,143,137,114,118,147,177,191,150,13
1,120,96,71,67,62,66,67,74,76,78,97,137,171,191,195,196,187,180,194,223,
225,228,226,209,195,191,192,179,176,167,166,170,188,205,227,227,231,23
5,236,230,220,228,233,230,229,218,205,195,192,183,180,176,164,161,155,
154,156,155,150,147,150,159,164,165,159,157,157,155,153,153,158,160,16
2,168,172,180,186,190,193,197,205,210,214,217,219,221,222,225,225,226,
227,226,228,232,230,224,223,219,214,216,215,213,216,216,212,209,209,20
9,207,207,204,204,205,204,204,204,198,195,191,187,187,186,182,167,148,
138,144,146,142,137,136,138,142,147,148,150,148,147,149,149,156,161,16
2,163,163,160,158,160,166,164,157,158,162,162,155,155,155,154,146,139

图3

段号	短语	编码
1	172	00010101100
2	32	00000100000
3	32,32	01000100000
4	54	00000110110
5	32,32,32	01100100000
6	54,172	10010101100
7	173	00010101101

图4

段号	短语	编码
1	172	00010101100
2	32,32,32	10000100000
3	54	00000110110
4	32,32,32,54	01000110110
5	172,173	00110101101

图5

0000000000000001011
0000000000000001010
0000000001000001010
0000000000000001001
0000000010000001001
0000000010100001001
0000000011000001001
0000000011100001001
0000000010000001000
0000000000000001000
0000000101000001000
0000000101100001000
0000000010100001010
0000000000100001011
0000000100100001000
0000000001000001011
0000000001100001101

图6

0000000000000001011
0000000010000001010
0000001001000001001
0000000101000001000
0000000011000001001
0000000000100001011
0000000000000001001
0000000100100001000
0000000000000001010
0000000000000001010
0000000100100001101

图7