



(12) 发明专利

(10) 授权公告号 CN 115276666 B

(45) 授权公告日 2022. 12. 20

(21) 申请号 202211188274.9

审查员 李硕

(22) 申请日 2022.09.28

(65) 同一申请的已公布的文献号

申请公布号 CN 115276666 A

(43) 申请公布日 2022.11.01

(73) 专利权人 汉达科技发展集团有限公司

地址 226000 江苏省南通市开发区华富国际大厦1幢816室

(72) 发明人 刘长宜 高国兵

(74) 专利代理机构 深圳信科专利代理事务所

(普通合伙) 44500

专利代理师 张琼

(51) Int. Cl.

H03M 7/30 (2006.01)

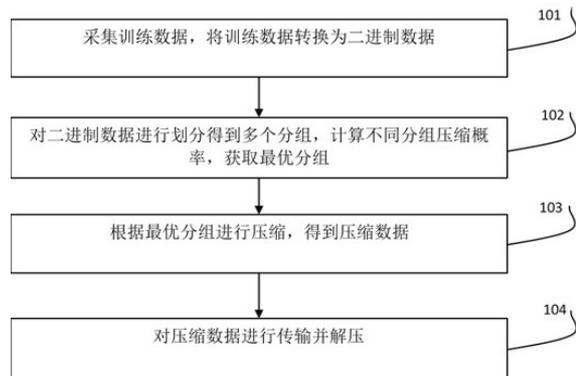
权利要求书2页 说明书8页 附图1页

(54) 发明名称

一种装备训练模拟器数据高效传输方法

(57) 摘要

本发明涉及数据传输技术领域,具体涉及一种装备训练模拟器数据高效传输方法,包括:采集训练数据,将训练数据转换为二进制数据;对二进制数据进行不同长度的划分,得到多个分组;根据每个分组的任意两个字符在二进制数据中相邻出现的次数获取字符串的第一次数,进一步构建每个分组的第一集合和第二集合,对第一集合和第二集合不断更新;根据每个分组的第一集合长度获取每个分组的压缩概率;将压缩概率最大的分组作为最优分组,对最优分组进行压缩。本发明通过对二进制数据进行划分,动态获取最优分组,压缩效率高,进而使得数据传输更加高效。



1. 一种装备训练模拟器数据高效传输方法,其特征在于,该方法包括以下步骤:

S1:采集训练数据,将训练数据转换为二进制数据;将二进制数据划分成多个分组,每个分组为多个字符的集合;

S2:获取每个分组的所有根,将任意两个根拼接成一个字符串,获取每个分组中每个字符串的第一次数,并得到每个分组的初始字典长度;每个分组中第一次数大于或等于0的所有字符串组成每个分组的第一集合;每个分组中第一次数大于等于1的所有字符串组成每个分组的第二集合;

S3:更新每个分组的第一集合与第二集合,包括:

将第二集合中第一次数最大的字符串作为第一字符串;获取第二集合中以第一字符串末尾字符为开头的字符串,得到多个第二字符串;根据第一字符串与每个第二字符串得到多个第三字符串;当存在第二字符串与第一字符串相同时,更新所述第二字符串以及第一字符串的第一次数;获取每个第二字符串的概率,根据第一字符串的第一次数以及每个第二字符串的概率获取每个第三字符串的第一次数;

将第一次数大于等于0的第三字符串添加到第一集合;将第一次数大于等于1的第三字符串添加到第二集合;更新第一字符串和每个第二字符串的第一次数;将第二集合中所有第一次数小于1的字符串从第二集合中删除;

S4:重复更新每个分组的第一集合与第二集合操作直到所述第二集合为空时停止;

S5:获取每个分组的第一集合长度,根据每个分组的第一集合长度以及每个分组的初始字典长度获得每个分组的压缩概率;获取压缩概率最大的分组作为最优分组,对最优分组进行编码压缩。

2. 根据权利要求1所述的一种装备训练模拟器数据高效传输方法,其特征在于,所述将二进制数据划分成多个分组,并得到每个分组的初始字典长度,每个分组为多个字符的集合,包括的具体步骤如下:

获取第一预设阈值与第二预设阈值之间所有的整数,构成分组长度集合;利用分组长度集合中的任意一个分组长度将二进制数据分成多个长度为所述分组长度的子二进制串,每个子二进制串视为一个字符,所有所述字符构成一个集合记为一个分组,将所述分组长度记为所述分组的分组长度;分组长度集合中的所有分组长度将二进制数据分成多个分组。

3. 根据权利要求1所述的一种装备训练模拟器数据高效传输方法,其特征在于,所述获取每个分组的所有根,将任意两个根拼接成一个字符串,获取每个分组中每个字符串的第一次数,并得到每个分组的初始字典长度,包括的具体步骤如下:

获取长度为每个分组的分组长度的所有二进制数,将每个所述二进制数视为一个字符,所有所述字符作为每个分组的多个根,将每个分组的根的个数作为每个分组的初始字典长度;

统计每个分组的任意两个根在所述分组中相邻出现的次数,将所述任意两个根拼接成一个字符串,将所述次数减一作为所述字符串的第一次数。

4. 根据权利要求1所述的一种装备训练模拟器数据高效传输方法,其特征在于,所述根据第一字符串与每个第二字符串得到多个第三字符串,包括的具体步骤如下:

将第一字符串与第二字符串根据第一字符串末尾字符进行拼接得到第三字符串,所述

第三字符串长度为第一字符串长度加第二字符串长度再减一,所述每个第二字符串对应一个第三字符串。

5.根据权利要求1所述的一种装备训练模拟器数据高效传输方法,其特征在于,所述当存在第二字符串与第一字符串相同时,更新所述第二字符串以及第一字符串的第一次数,包括的具体步骤如下:

当存在第二字符串与第一字符串相同时,将第一字符串的第一次数除以所有第二字符串的个数再向上取整,结果作为所述第二字符串的第一次数;将第一字符串的第一次数减去所述第二字符串的第一次数,所得结果作为第一字符串的第一次数。

6.根据权利要求1所述的一种装备训练模拟器数据高效传输方法,其特征在于,所述获取每个第二字符串的概率,根据第一字符串的第一次数以及每个第二字符串的概率获取每个第三字符串的第一次数,包括的具体步骤如下:

计算每个第二字符串的第一次数占所有第二字符串第一次数总和的比例,将所述比例作为所述第二字符串的概率;

将第一字符串的第一次数乘以每个第二字符串的概率并四舍五入取整再减一得到每个第三字符串的第一次数。

7.根据权利要求1所述的一种装备训练模拟器数据高效传输方法,其特征在于,所述更新第一字符串和每个第二字符串的第一次数,包括的具体步骤如下:

将第一字符串的第一次数更新为0;将每个第二字符串的第一次数减去所述第二字符串对应的第三字符串的第一次数,作为所述第二字符串更新后的第一次数;当存在第二字符串与第一字符串相同时,将第一字符串的第一次数更新为所述第二字符串更新后的第一次数。

8.根据权利要求1所述的一种装备训练模拟器数据高效传输方法,其特征在于,所述获取每个分组的第一集合长度,根据每个分组的第一集合长度以及每个分组的初始字典长度获得每个分组的压缩概率,包括的具体步骤如下:

将第一集合中的元素个数作为第一集合长度,将第一集合长度减一得到预测代码个数;将第一集合长度与初始字典长度相加得到预测最大代码值;根据预测代码个数与预测最大代码值以及二进制数据的长度获得分组的压缩概率;所述压缩概率的表达式为:

$$q = 1 - \frac{s \times \lceil \log_2 t \rceil}{d}$$

其中 $q$ 为压缩概率; $s$ 为预测代码个数; $t$ 为预测最大代码值; $d$ 为二进制数据的长度; $\lceil \cdot \rceil$ 为向上取整符。

9.根据权利要求1所述的一种装备训练模拟器数据高效传输方法,其特征在于,所述对最优分组进行编码压缩是指对最优分组利用LZW编码进行压缩。

## 一种装备训练模拟器数据高效传输方法

### 技术领域

[0001] 本发明涉及数据传输技术领域,具体涉及一种装备训练模拟器数据高效传输方法。

### 背景技术

[0002] 装备训练模拟器是利用仿真技术对人员进行训练的自动化设备,广泛应用于航天、航空、军事、医学等领域。

[0003] 装备训练模拟器在对人员进行训练过程中会产生大量训练数据。在训练结束后,需要将训练数据传输至智能分析平台。智能分析平台依据训练数据对训练结果进行评价。为确保训练数据的高效传输,需要对训练数据进行压缩。

[0004] 装备训练模拟器在训练过程中产生的训练数据包含多种类型的数据。不同类型的数据精度不同,且数据重复率非常小。

[0005] 现有的压缩算法如LZW、霍夫曼编码、游程编码等对于重复率大的数据具有较好的压缩效果,对于重复率小的数据压缩效果差。现有的压缩算法如LZW、霍夫曼编码、游程编码不适用于对训练数据进行压缩。

### 发明内容

[0006] 本发明提供一种装备训练模拟器数据高效传输方法,以解决现有的问题。

[0007] 本发明的一种装备训练模拟器数据高效传输方法采用如下技术方案:

[0008] 本发明一个实施例提供了一种装备训练模拟器数据高效传输方法,该方法包括以下步骤:

[0009] S1:采集训练数据,将训练数据转换为二进制数据;将二进制数据划分成多个分组,每个分组为多个字符的集合;

[0010] S2:获取每个分组的所有根,将任意两个根拼接成一个字符串,获取每个分组中每个字符串的第一次数,并得到每个分组的初始字典长度;每个分组中第一次数大于或等于0的所有字符串组成每个分组的第一集合;每个分组中第一次数大于等于1的所有字符串组成每个分组的第二集合;

[0011] S3:更新每个分组的第一集合与第二集合,包括:

[0012] 将第二集合中第一次数最大的字符串作为第一字符串;获取第二集合中以第一字符串末尾字符为开头的字符串,得到多个第二字符串;根据第一字符串与每个第二字符串得到多个第三字符串;当存在第二字符串与第一字符串相同时,更新所述第二字符串以及第一字符串的第一次数;获取每个第二字符串的概率,根据第一字符串的第一次数以及每个第二字符串的概率获取每个第三字符串的第一次数;

[0013] 将第一次数大于等于0的第三字符串添加到第一集合;将第一次数大于等于1的第三字符串添加到第二集合;更新第一字符串和每个第二字符串的第一次数;将第二集合中所有第一次数小于1的字符串从第二集合中删除;

[0014] S4:重复更新每个分组的第一集合与第二集合操作直到所述第二集合为空时停止;

[0015] S5:获取每个分组的第一集合长度,根据每个分组的第一集合长度以及每个分组的初始字典长度获得每个分组的压缩概率;获取压缩概率最大的分组作为最优分组,对最优分组进行编码压缩。

[0016] 优选的,所述将二进制数据划分成多个分组,并得到每个分组的初始字典长度,每个分组为多个字符的集合,包括的具体步骤如下:

[0017] 获取第一预设阈值与第二预设阈值之间所有的整数,构成分组长度集合;利用分组长度集合中的任意一个分组长度将二进制数据分成多个长度为所述分组长度的子二进制串,每个子二进制串视为一个字符,所有所述字符构成一个集合记为一个分组,将所述分组长度记为所述分组的分组长度;分组长度集合中的所有分组长度将二进制数据分成多个分组。

[0018] 优选的,所述获取每个分组的所有根,将任意两个根拼接成一个字符串,获取每个分组中每个字符串的第一次数,并得到每个分组的初始字典长度,包括的具体步骤如下:

[0019] 获取长度为每个分组的分组长度的所有二进制数,将每个所述二进制数视为一个字符,所有所述字符作为每个分组的多个根,将每个分组的根的个数作为每个分组的初始字典长度;

[0020] 统计每个分组的任意两个根在所述分组中相邻出现的次数,将所述任意两个根拼接成一个字符串,将所述次数减一作为所述字符串的第一次数。

[0021] 优选的,所述根据第一字符串与每个第二字符串得到多个第三字符串,包括的具体步骤如下:

[0022] 将第一字符串与第二字符串根据第一字符串末尾字符进行拼接得到第三字符串,所述第三字符串长度为第一字符串长度加第二字符串长度再减一,所述每个第二字符串对应一个第三字符串。

[0023] 优选的,所述当存在第二字符串与第一字符串相同时,更新所述第二字符串以及第一字符串的第一次数,包括的具体步骤如下:

[0024] 当存在第二字符串与第一字符串相同时,将第一字符串的第一次数除以所有第二字符串的个数再向上取整,结果作为所述第二字符串的第一次数;将第一字符串的第一次数减去所述第二字符串的第一次数,所得结果作为第一字符串的第一次数。

[0025] 优选的,所述获取每个第二字符串的概率,根据第一字符串的第一次数以及每个第二字符串的概率获取每个第三字符串的第一次数,包括的具体步骤如下:

[0026] 计算每个第二字符串的第一次数占所有第二字符串第一次数总和的比例,将所述比例作为所述第二字符串的概率;

[0027] 将第一字符串的第一次数乘以每个第二字符串的概率并四舍五入取整再减一得到每个第三字符串的第一次数。

[0028] 优选的,所述更新第一字符串和每个第二字符串的第一次数,包括的具体步骤如下:

[0029] 将第一字符串的第一次数更新为0;将每个第二字符串的第一次数减去所述第二字符串对应的第三字符串的第一次数,作为所述第二字符串更新后的第一次数;当存在第

二字符串与第一字符串相同时,将第一字符串的第一次数更新为所述第二字符串更新后的第一次数。

[0030] 优选的,所述获取每个分组的第一集合长度,根据每个分组的第一集合长度以及每个分组的初始字典长度获得每个分组的压缩概率,包括的具体步骤如下:

[0031] 将第一集合中的元素个数作为第一集合长度,将第一集合长度减一得到预测代码个数;将第一集合长度与初始字典长度相加得到预测最大代码值;根据预测代码个数与预测最大代码值以及二进制数据的长度获得分组的压缩概率;所述压缩概率的表达式为:

$$[0032] \quad q = 1 - \frac{s \times \lceil \log_2 t \rceil}{d}$$

[0033] 其中 $q$ 为压缩概率; $s$ 为预测代码个数; $t$ 为预测最大代码值; $d$ 为二进制数据的长度; $\lceil \cdot \rceil$ 为向上取整符。

[0034] 优选的,所述对最优分组进行编码压缩是指对最优分组利用LZW编码进行压缩。

[0035] 本发明的有益效果是:采集训练数据,将训练数据转换为二进制数据;对二进制数据进行不同长度的划分,得到多个分组;根据每个分组的任意两个字符在二进制数据中相邻出现的次数获取字符串的第一次数,进一步构建每个分组的第一集合和第二集合,对第一集合和第二集合不断更新;根据每个分组的第一集合长度获取每个分组的压缩概率;将压缩概率最大的分组作为最优分组,对最优分组进行压缩。本发明通过将训练数据转换为二进制数据并对二进制数据进行不同长度的划分,一定程度增加了数据的重复率,使得训练数据压缩效率更高。根据每个分组中字符出现的规律构建第一集合,进一步计算压缩概率,可一定程度预测出不同分组的压缩效率,从而获取最优的分组,动态实现当前方法下的最高压缩效率。使得数据传输更加高效。

## 附图说明

[0036] 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。

[0037] 图1为本发明的一种装备训练模拟器数据高效传输方法的步骤流程图。

## 具体实施方式

[0038] 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装备训练模拟器数据高效传输方法,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。

[0039] 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。

[0040] 下面结合附图具体的说明本发明所提供的一种装备训练模拟器数据高效传输方

法的具体方案。

[0041] 请参阅图1,其示出了本发明一个实施例提供的一种装备训练模拟器数据高效传输方法的步骤流程图,该方法包括以下步骤:

[0042] 101.采集训练数据,将训练数据转换为二进制数据。

[0043] 装备训练模拟器在对人员进行训练过程中会产生大量训练数据。在训练结束后,需要将训练数据传输至智能分析平台。智能分析平台依据训练数据对训练结果进行评价。为确保训练数据的高效传输,需要对训练数据进行压缩。首先采集训练数据。

[0044] 训练数据包含多种不同类型的数据,不同类型的数据精度不同,且数据重复率非常小。现有的压缩算法如LZW、霍夫曼编码、游程编码等对于重复率大的数据具有较好的压缩效果,对于重复率小的数据压缩效果差。若直接采用现有压缩方法对训练数据进行压缩,压缩效果差,不利于训练数据的高效传输。

[0045] 无论是哪种类型的数据,在进行存储传输时都需要转换为二进制数据。因此,将训练数据转换为二进制数据,二进制数据的长度记为 $d$ 。本发明实施例针对二进制数据进行动态自适应分组压缩。

[0046] 102.对二进制数据进行划分得到多个分组,计算不同分组压缩概率,获取最优分组。

[0047] 具体步骤如下:

[0048] 1、获取分组长度集合。

[0049] 本发明实施例的目的是将二进制数据分成多个子二进制串,使得多个子二进制串之间重复率高,从而利用LZW编码可以达到较好的压缩效果。

[0050] 当分组长度越小时,得到的子二进制串的个数较多,子二进制串的重复率较大;当分组长度越大时,得到的子二进制串的个数少,子二进制串的重复率较小。

[0051] 利用LZW编码对所有子二进制串进行压缩,压缩效率与子二进制串的个数以及子二进制串的重复率有关。当子二进制串的个数越少且重复率越大时,压缩效率较高;当子二进制串的个数越多且重复率越小时,压缩效率较低。因此,需获得一个最优的分组长度,使得子二进制串的个数少且重复率大。

[0052] 首先构建一个分组长度集合:当分组长度等于 $a = 1$ 时,分组无意义;当分组长度过大时,不同的子二进制串个数非常多,导致所有子二进制串的重复率非常小,从而使得压缩效率非常差。设置一个最大分组长度 $b = 10$ ,当分组长度为 $10$ 时,存在 $2^{10}$ 种不同的子二进制串,此时所有子二进制串的重复率非常小。因此,将区间 $(a, b)$ 作为分组长度的范围,获取

区间 $(a, b)$ 内的所有整数,构成分组长度集合。在其他实施例中,实施人员可根据需要设置分组长度集合。

[0053] 2、对二进制数据进行分组。

[0054] 利用分组长度集合中的一个分组长度将二进制数据分成多个长度为该分组长度的子二进制串,将每个子二进制串视为一个字符。将所有字符构成一个集合,作为一个分组,将该分组长度作为该分组的分组长度。

[0055] 同理,利用分组长度集合中的每个分组长度对二进制数据进行划分,得到多个分组。

[0056] 若直接对不同的分组进行压缩,可根据压缩结果获取最优的分组。但在压缩过程中,需要动态构建字典,若对所有分组都进行压缩,则会构建多个字典,消耗大量的内存空间,同时时间效率差。

[0057] 因此,本发明实施例通过分析不同分组中子二进制串的分布规律,计算不同分组的概率从而选择最优的分组。

[0058] 3、计算每个分组的压缩概率。

[0059] 本发明实施例的目的是对二进制数据分组后,再利用LZW编码可以达到较好的压缩效果。而LZW编码是在初始字典的基础上不断地为数据中新出现的字符串赋予代码并添加到字典中,用前面已经出现的字符串的代码来代替后面相同字符串的内容,从而达到压缩的目的。为确保数据中新出现的每个单个字符都可用代码表示,即初始字典中需包含所有可能的单个字符。

[0060] LZW编码的具体步骤如下:

[0061] ①构建初始字典,初始字典需包含所有可能的单个字符,单个字符称为根。构建一个空的前缀P。

[0062] ②读入新的字符C,与前缀P合并形成字符串P+C。

[0063] ③在字典中查找字符串P+C:

[0064] a) 如果字符串P+C在字典里,设置前缀P=P+C;

[0065] b) 如果P+C不在字典里,将前缀P在字典中的代码输出。将P+C添加到字典中。更新前缀P=C。

[0066] ④返回步骤②重复,直到读完数据中所有的字符。

[0067] 在LZW编码压缩结束后,输出的为代码流。LZW的压缩效率与输出的代码流中代码的个数以及每个代码的编码长度相关。

[0068] 本发明实施例分组后每个子二进制串可以看作单个字符,即每个子二进制串可以看作一个根。本发明实施例使用十进制数字作为代码。而数据传输或存储都需要转换成二进制形式,因此需将最终压缩得到的代码流中的每个代码转换为长度相同的二进制串,此些二进制串作为每个代码的编码。每个代码的编码长度与最大的代码值相关,假设最大的代码值为 $X$ ,则每个代码的编码长度为 $\lceil \log_2 X \rceil$ ,其中 $\lceil \cdot \rceil$ 为向上取整符。最大代码值与字典的长度有关。

[0069] 在LZW编码过程中,除最后一个代码输出时不添加字典外,每输出一代码,就需要往字典中添加一次内容。因此输出的代码流中代码的个数与字典长度有关,输出的代码流中代码的个数为字典添加内容的次数加一,即最终的字典的长度减去初始字典长度再加一。

[0070] 本发明根据分组后所有子二进制串规律预测字典的长度,进一步获取预测最大代码值以及输出的代码流中代码的预测个数,以便计算每个分组的压缩概率从而选择最优的分组。将输出的代码流中代码的预测个数称为预测代码个数。

[0071] 以一个分组为例,说明字典长度的预测方法,获取预测最大代码值以及预测代码

个数,计算分组的压缩概率。具体如下:

[0072] 将分组的分组长度记为 $k$ ,则分组中的每个子二进制串的长度为 $k$ 。长度为 $k$ 的二进制数有 $2^k$ 种,即根有 $2^k$ 种。初始字典需包含所有可能的根,因此初始字典长度为 $2^k$ 。

[0073] 统计所有根中任意两个根在分组中相邻出现的次数,得到根共生矩阵。根共生矩阵中的值表示对应行的根与对应列的根在分组中相邻出现的次数,将任意两个根拼接得到一个字符串,即根共生矩阵中的值表示对应字符串在分组中出现的次数。假设二进制数据101100010的一个分组为101,100,010,则101与100相邻,100与010相邻。

[0074] 根据LZW编码的压缩原理可知,当字符串第一次出现时,该字符串会被添加到字典中,当字符串第二次出现时,会将字符串向后扩充一个字符串添加到字典中。为获取所有在字典中可能出现的字符串,将每个字符串在分组中相邻出现的次数减1记为第一次数。

[0075] 构建一个空的第一集合用来存放所有可能添加到字典中的字符串。构建一个空的第二集合用来存放预测字典的长度过程中产生的字符串。根据LZW编码的压缩原理可知,当字符串第一次出现时,需往字典中添加该字符串。因此,将所有第一次数大于或等于0的字符串添加到第一集合中;当字符串第二次出现时,则会输出该字符串在字典中的代码,同时会将字符串向后扩充一个字符串添加到字典中。将所有第一次数大于或等于1的字符串添加到第二集合中。

[0076] 进行第一集合与第二集合更新操作:

[0077] 获取测最长字符串集合中第一次数最大的字符串,记作第一字符串,第一字符串的长度记为 $N1$ ,第一字符串在根公式矩阵记为 $n$ ,将第一字符串末尾的字符记作第一字符。获取第二集合中所有以第一字符为开头的字符串,记作第二字符串,第二字符串的长度记为 $N2$ ,所有第二字符串的个数记为 $m$ 。

[0078] 第一字符串与一个第二字符串可组成一个长度 $N1 + N2 - 1$ 的字符串,将该字符串记为第三字符串,例如第一字符串AB与第二字符串BC构成第三字符串ABC。获取所有的第三字符串。预测每个第三字符串出现的次数:

[0079] 可能存在第二字符串与第一字符串相同的情况,例如第一字符串AA与第二字符串AA构成第三字符串AAA。当一个第二字符串与第一字符串相同时,该第二字符串与第一字符串实际为同一个字符串,为避免该字符串被重复计算,将该第二字符串的第一次数调整为

$\lceil \frac{n}{m} \rceil$ ,其中 $\lceil \cdot \rceil$ 为向上取整符;将第一字符串的第一次数调整为 $n - \lceil \frac{n}{m} \rceil$ 。当不存在第二字符串与第一字符串相同时,则不调整第二字符串与第一字符串的第一次数。

[0080] 计算每个第二字符串的第一次数占所有第二字符串的第一次数之和的比例,得到每个第二字符串的概率。则每个第三字符串预计出现的次数为第一字符串的第一次数乘以组成第三字符串的第二字符串的概率并四舍五入取整。

[0081] 将每个第三字符串预计出现的次数减一,得到每个第三字符串的第一次数。将所有第一次数大于或等于0的第三字符串添加到第一集合中。将所有第一次数大于或等于1的第三字符串添加到第二集合中。由于第一字符串已经分配给所有第二字符串,因此将第一字符串的第一次数更新为0。由于每个第二字符串与第一字符串构成了第三字符串,因此根

据第三字符串预计出现的次数更新第二字符串的第一次数,将第二字符串的第一次数减去其构成的第三字符串的预计出现的次数,所得结果作为第二字符串更新后的第一次数。当存在第二字符串与第一字符串相同时,将第一字符串的第一次数更新为所述第二字符串更新后的第一次数;将第二集合中所有第一次数小于1的字符串删除。

[0082] 重复第一集合与第二集合更新操作直到第二集合为空时停止。

[0083] 至此,得到了第一集合,获取了所有可能添加到字典中的字符串。获取第一集合中元素的个数 $l$ ,将其作为第一集合长度,则分组的字典长度预测值为 $t$ :

$$[0084] \quad t = 2^k + l$$

[0085] 其中 $2^k$ 为初始字典长度; $l$ 为第一集合长度。

[0086] 则预测代码个数为 $s = l - 1$ 。最大代码值为输出的代码流中最大的代码,最大代码值与字典长度有关。可将字典长度预测值近似看作最大代码值,则预测最大代码值为 $t$ 。

[0087] 根据预测最大代码值以及预测代码个数计算分组的压缩概率 $q$ :

$$[0088] \quad q = 1 - \frac{s \times \lceil \log_2 t \rceil}{d}$$

[0089] 其中 $s$ 为预测代码个数; $t$ 为预测最大代码值; $d$ 为二进制数据的长度; $\lceil \log_2 t \rceil$ 表示每个代码的编码长度;当压缩概率越大时,压缩效果越好。当压缩概率越小时,压缩效果越差。

[0090] 同理,计算每个分组的压缩概率。

[0091] 4、获取最优分组。

[0092] 将压缩概率最大的分组作为最优分组。

[0093] 至此,获取了最优分组。

[0094] 103. 根据最优分组进行压缩,得到压缩数据。

[0095] 将最优分组对应的分组长度作为最优分组长度,根据最优分组长度获取所有可能的根,建立初始字典:假设最优分组长度为 $k$ ,则长度为 $k$ 的所有可能的二进制数有 $2^k$ 种,此些二进制数即为可能的根。而LZW编码的初始字典需包含所有可能的根,因此将 $2^k$ 种长度为 $k$ 的二进制数的十进制值作为此些二进制数的代码,并将此些二进制数与其对应的代码按照从小到大的顺序放入到字典中,得到初始字典。

[0096] 再根据初始字典利用LZW编码的压缩方法对最优分组进行压缩,得到压缩数据。

[0097] 104. 对压缩数据进行传输并解压。

[0098] 装备训练模拟器将压缩数据以及最优分组长度传输至智能分析平台。智能分析平台根据最优分组长度获取所有可能的根,建立初始字典。再根据初始字典利用LZW解码的方法对压缩数据进行解压,得到多个子二进制串。

[0099] 将所有子二进制串按照顺序拼接在一起得到压缩前的二进制数据。将二进制数据转换为训练数据。

[0100] 至此,完成了训练数据的解压,智能分析平台根据训练数据对训练结果进行评价。

[0101] 通过以上步骤,完成了装备训练模拟器训练数据的压缩以及高效传输。

[0102] 本发明实施例通过采集训练数据,将训练数据转换为二进制数据;对二进制数据进行不同长度的划分,得到多个分组;根据每个分组的任意两个字符在二进制数据中相邻出现的次数获取字符串的第一次数,进一步构建每个分组的第一集合和第二集合,对第一集合和第二集合不断更新;根据每个分组的第一集合长度获取每个分组的压缩概率;将压缩概率最大的分组作为最优分组,对最优分组进行压缩。本发明通过将训练数据转换为二进制数据并对二进制数据进行不同长度的划分,一定程度增加了数据的重复率,使得训练数据压缩效率更高。根据每个分组中字符出现的规律构建第一集合,进一步计算压缩概率,可一定程度预测出不同分组的压缩效率,从而获取最优的分组,动态实现当前方法下的最高压缩效率。使得数据传输更加高效。

[0103] 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。

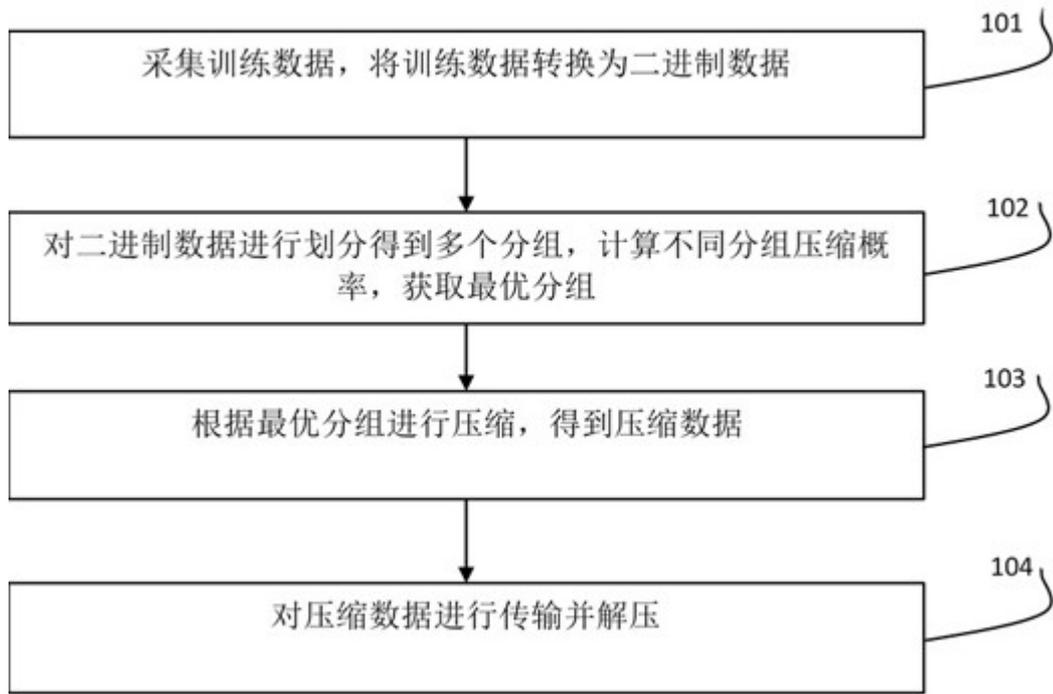


图1