(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
2017 年 7 月 27 日 (27.07.2017)

WIPO | PCT

(10) 国际公布号
WO 2017/124974 A1

(51) 国际专利分类号:
G06F 17/18 (2006.01)

(21) 国际申请号: PCT/CN2017/071130

(22) 国际申请日: 2017 年 1 月 13 日 (13.01.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201610045655.X 2016 年 1 月 22 日 (22.01.2016) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。

(72) 发明人; 及

(71) 申请人 (仅对美国): 周俊 (ZHOU, Jun) [CN/CN]; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北京国昊天诚知识产权代理有限公司 (CO-HORIZON INTELLECTUAL PROPERTY INC.); 中国北京市朝阳区小关北里甲 2 号渔阳置业大厦 B 座 605, Beijing 100029 (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, 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, 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,

[见续页]

(54) Title: CALCULATION METHOD AND DEVICE FOR LOGICAL REGRESSION GRADIENT

(54) 发明名称 : 逻辑回归梯度的计算方法和装置

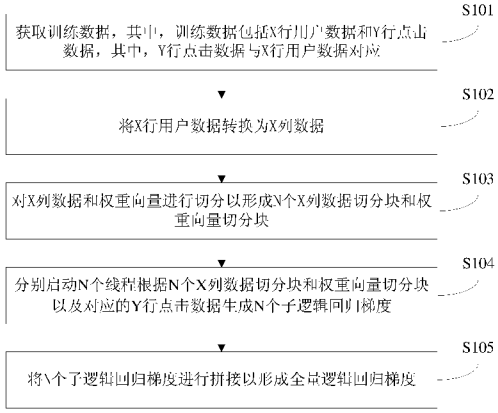


图 1

S101 Acquire training data, where the training data comprises X rows of user data and Y rows of click data, and the Y rows of click data correspond to the X rows of user data

S102 Convert the X rows of user data into X columns of data

S103 Divide the X columns of data and a weight vector to form N divided blocks of the X columns of data and divided blocks of the weight vector

S104 Respectively start N threads to generate N logical regression sub-gradients on the basis of the N divided blocks of the X columns of data, the divided blocks of the weight vector, and the corresponding Y rows of click data

S105 Splice the N logical regression sub-gradients to form a full logical regression gradient

(57) **Abstract**: A calculation method and device for a logical regression gradient. The calculation method comprises: acquiring training data, where the training data comprises X rows of user data and Y rows of click data, and the Y rows of click data correspond to the X rows of user data (S101); converting the X rows of user data into X columns of data (S102); dividing the X columns of data and a weight vector so as to form N divided blocks of the X columns of data and divided blocks of the weight vector (S103); respectively starting N threads to generate N logical regression sub-gradients on the basis of the N divided blocks of the X rows of data and divided blocks of the weight vector and the corresponding Y rows of click data (S104); and splicing the N logical regression sub-gradients to form a full logical regression gradient (S105). The calculation method allows a computer to support the training of an ultra-large-scale logical regression model, increases calculation speed, reduces training time, and greatly reduces the amount of memory used on the computer.

(57) **摘要**: 一种逻辑回归梯度的计算方法和装置。该计算方法包括：获取训练数据，其中，训练数据包括 X 行用户数据和 Y 行点击数据，其中，Y 行点击数据与 X 行用户数据对应（S101）；将 X 行用户数据转换为 X 列数据（S102）；对 X 列数据和权重向量进行切分以形成 N 个 X 列数据切分块和权重向量切分块（S103）；分别启动 N 个线程根据 N 个 X 列数据切分块和权重向量切分块以及对应的 Y 行点击数据生成 N 个子逻辑回归梯度（S104）；以及将 N 个子逻辑回归梯度进行拼接以形成全量逻辑回归梯度（S105）。该计算方法使得计算机器可以支持超大规模逻辑回归模型的训练，提高计算速度，减少训练时间，并且极大地降低计算机器的内存使用量。

WO 2017/124974 A1

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))。

逻辑回归梯度的计算方法和装置

技术领域

本申请涉及互联网技术领域，尤其涉及一种逻辑回归梯度的计算方法和装置。

背景技术

逻辑回归作为网络广告 CTR（Click-Through-Rate，点击到达率）中最为常用的模型，在业界中被大规模使用。谁能在有限的时间里面训练出逻辑回归模型，就能更好地为用户提供合理的广告，提升用户的使用体验。

在训练逻辑回归模型的过程中，最为核心的部分是梯度的计算。现有的梯度的计算方法，是通过在计算机器上启动多线程的方式来进行加速，然而，每个线程均需要一份全量的梯度向量存储在内存中。也就是说，在一台计算机器上如果启动 10 个线程来计算梯度，假设 1 份梯度向量为 10 亿，均为 double 数据，那么 1 个线程需要消耗约 7.5G 的内存，一共需要消耗计算机器 75G 的内存。这对于现有的计算机器来说通常是很难承受的。

进一步而言，在互联网领域中，要想提高网络广告的 CTR，就必须要使用更多的训练数据，更大的特征来刻画长尾，从而满足用户需求。例如，在训练逻辑回归模型时,考虑到用户的性别、年龄、目前所在地等细粒度的特征，能够更加有效刻画用户的需求，从而更加精准的满足用户个性化需求。因此，如此大的逻辑回归模型使用现有技术很难进行训练。

申请内容

本申请旨在至少在一定程度上解决相关技术中的技术问题之一。

为此，本申请的第一个目的在于提出一种逻辑回归梯度的计算方法，该计算方法使得计算机器可以支持超大规模逻辑回归模型的训练，提高计算速度，减少训练时间，并且极大地降低计算机器的内存使用量。

本申请的第二个目的在于提出一种逻辑回归梯度的计算装置。

为达上述目的，本申请第一方面实施例提出了一种逻辑回归梯度的计算方法，包括以下步骤：获取训练数据，其中，所述训练数据包括X行用户数据和Y行点击数据，其中，所述Y行点击数据与所述X行用户数据对应；将所述X行用户数据转换为X列数据；对所述X列数据和权重向量进行切分以形成N个X列数据切分块和权重向量切分块；分别启动N个线程根据所述N个X列数据切分块和权重向量切分块以及对应的所述Y行点击数据生成N个子逻辑回归梯度；以及将所述N个子逻辑回归梯度进行拼接以形成全量逻辑回归梯度。

本申请实施例的计算方法，将训练数据和权重向量切分成内存中可以存放下的多个小区域，并启动对应个数的线程对个小区域进行并行计算，由此使得计算机器可以支持超大规模逻辑回归模型的训练，提高计算速度，减少训练时间，并且极大地降低计算机器的内存使用量。

为达上述目的，本申请第二方面实施例提出了一种逻辑回归梯度的计算装置，包括：获取模块，用于获取训练数据，其中，所述训练数据包括X行用户数据和Y行点击数据，其中，所述Y行点击数据与所述X行用户数据对应；转换模块，用于将所述X行用户数据转换为X列数据；切分模块，用于对所述X列数据和权重向量进行切分以形成N个X列数据切分块和权重向量切分块；生成模块，用于分别启动N个线程根据所述N个X列数据切分块和权重向量切分块以及对应的所述Y行点击数据生成N个子逻辑回归梯度；拼接模块，用于将所述N个子逻辑回归梯度进行拼接以形成全量逻辑回归梯度。

本申请实施例的计算装置，将训练数据和权重向量切分成内存中可以存放下的多个小区域，并启动对应个数的线程对个小区域进行并行计算，由此使得计算机器可以支持超大规模逻辑回归模型的训练，提高计算速度，减少训练时间，并且极大地降低计算机器的内存使用量。

本申请附加的方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本申请的实践了解到。

附图说明

本申请上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解，其中：

图 1 是本申请一个实施例的逻辑回归梯度的计算方法的流程图；

图 2 是本申请一个实施例的通过列切分装置将权重向量 W 进行切分的示意图；

图 3 是本申请另一个实施例的逻辑回归梯度的计算方法的流程图；

图 4 是本申请一个实施例的逻辑回归梯度的计算装置的结构示意图；以及

图 5 是本申请另一个实施例的逻辑回归梯度的计算装置的结构示意图。

具体实施方式

下面详细描述本申请的实施例，所述实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，旨在用于解释本申请，而不能理解为对本申请的限制。

此外，术语“第一”、“第二”仅用于描述目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个该特征。在本申请的描述中，“多个”的含义是两个或两个以上，除非另有明确具体的限定。

流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为，表示包括一个或更多个用于实现特定逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分，并且本申请的优选实施方式的范围包括另外的实现，其中可以不按所示出或讨论的顺序，包括根据所涉及的功能按基本同时的方式或按相反的顺序，来执行功能，这应被本申请的实施例所

属技术领域的技术人员所理解。

具体而言，逻辑回归计算公式为：

$$P(y|x,w) = 1/(1+\exp(-wx)),$$

其中， w 为权重向量，x 为训练数据。通过上述的公式来预测用户的点击概率。因此，在通过梯度下降算法来训练逻辑回归模型时，梯度的计算公式为：

$$1/(1+\exp(ywx)) * y * x,$$

其中，y 为用户的点击结果，由于我们拿到的日志是用户已经操作后的，如果用户进行了点击，则$y = 1$；如果用户没有点击，则$y = -1$。

目前，相关技术中是启动多个线程，每个线程均有一个跟权重向量 w 一样大的梯度向量 g，其中，第 i 个坐标的梯度为：

$$g_i = \sum_{s1}^{sn}(1/(1+\exp(ywx) * y * x_i),$$

其中，s1为第一个训练数据，sn为第 n 个训练数据。每个线程均遍历属于该线程的所有的训练数据，从而得到一个完整的梯度向量g_i。然后，合并所有线程的计算结果，得到最终的梯度向量g。

然而，目前存在的缺点是，当特征的规模很大时，每个计算机器需要的内存为线程个数*权重向量 w，导致计算机器的内存消耗巨大，单个计算机器通常加载不了，无法训练逻辑回归模型。此外，无法利用更多的 CPU 进行计算，因为 CPU 用的越多，计算机器的内存消耗的就越大，导致训练速度慢，训练成本大。

然而，本申请中逻辑回归梯度的计算方法将权重空间和训练数据切分成内存中可以存放下的多个更小的区域，并将该多个区域并行计算，由此可以提高计算速度，减少了训练时间，并且极大地降低了计算机器的内存使用量。

进一步而言，在计算梯度时，将权重向量和训练数据按列进行切分，每个线程计算一部分切分后的权重列和对应的数据列。由此，每个线程只需要持有梯度的不同区域，因此计算梯度所消耗的计算机器的内存由原来

的 N 份（其中，计算机器使用 N 个 CPU，就消耗 N 份计算机器的内存，例如，计算机器使用 10 个 CPU 来参与梯度计算，那么 N=10）变成 1 份，由此极大的降低了计算机器的内存消耗。而在计算速度上，原来最多只能启动几个线程，现在使用了多少个 CPU，就可以启动多少个线程，不受限于计算机器的内存限制，由此提高了计算机器的 CPU 利用率。因此，通过这样的方式，可以完成超大规模逻辑回归模型的训练，同时训练逻辑回归模型时不受限于计算机器的内存和 CPU 的限制，可以使用更多的资源来加速训练，由此降低了训练的时间跟成本。也就是说，本申请中的逻辑回归梯度的计算方法，不仅可以降低逻辑回归模型的训练成本，还可以实现训练超大规模的逻辑回归模型，从而能够更加有效的刻画长尾需求，满足用户的个性化需求。

图 1 是本申请一个实施例的逻辑回归梯度的计算方法的流程图。

如图 1 所示，计算方法包括:

S101，获取训练数据，其中，训练数据包括 X 行用户数据和 Y 行点击数据，其中，Y 行点击数据与 X 行用户数据对应。

例如，以以下格式先输入用户数据:

X:[1.3.5.8.9.10.23.22

2.4.6.8.9.12.34.56

.....

1.3.6.8.10.11.12],

其中，第一行为第一个用户数据 X，其中 1.3.5.8.9.10.23.22 代表第一个用户数据 X 中的不同用户特征；第二行为第二个用户数据 X，其中 2.4.6.8.9.12.34.56 代表第二个用户数据 X 中的不同用户特征；而第 N 行为第 N 个用户数据，其中 1.3.6.8.10.11.12 代表第 N 个用户数据中的不同用户特征。

进而，再以以下格式输入点击数据:

Y:[1.0.0.0

2.0.0.0

……

0.0.1.0],

其中，第一行为第一个点击数据 Y；第二行为第二个点击数据 Y；而第 N 行为第 N 个点击数据。

应当理解的是，上述第一行的第一个用户数据 X 和第一行的第一个点击数据 Y 合并在一起，才是第一个完整的训练数据；第二行的第二个用户数据 X 和第二行的第二个点击数据 Y 合并在一起，才是第二个完整的训练数据；而第 N 行的第 N 个用户数据 X 和第 N 行的第 N 个点击数据 Y 合并在一起，才是第 N 个完整的训练数据。

S102，将 X 行用户数据转换为 X 列数据。

具体地，通过列数据转换装置将 X 行用户数据转换为 X 列数据。

S103，对 X 列数据和权重向量进行切分以形成 N 个 X 列数据切分块和权重向量切分块。

具体地，通过列切分装置将 X 列数据进行切分以得到对应的 N 个 X 列数据切分块。例如，N=3，即将 X 列数据切分为 3 个 X 列数据切分块，这三个 X 列数据切分块分别是：

X1: [1.3.5	X2: [8.9	X3: [10.23.22
2.4.6	8.9	12.34.56
……	……	……
1.3.6]	8]	10.11.12]

进而，如图 2 所示，通过列切分装置将权重向量 W 进行切分以得到对应的权重向量切分块。例如，W1 对应 W 中下标为 1-7 的元素，W2 对应 W 中下标为 8-9 的元素，而 W3 对应 W 中下表为 10-56 的元素。

在本发明的一个实施例中，X 行用户数据中的列数与权重向量的向量 W 宽度相同。也就是说，X 行用户数据的列数和权重向量 W 一样宽，例如，如果权重向量 W 的长度是 100，那么 X 行用户数据中每行用户数据中

最大出现的元素也为 100，因此 X 行用户数据中的每行用户数据都是稀疏的，也就是说每个用户数据只包括一部分的特征。例如，湖南的男性用户跟北京的女性用户，前者出现了湖南和男性的特征，后者出现了北京和女性的特征，对应到 X 行用户数据中的两行用户数据中，第一行可能是 1.7，而第二行可能是 9.10。X 行用户数据中 Y 也有两列，第一列 Y 可能是点击，第二列 Y 可能是未点击。

S104，分别启动 N 个线程根据 N 个 X 列数据切分块和权重向量切分块以及对应的 Y 行点击数据生成 N 个子逻辑回归梯度。

在本发明的一个实施例中， N 为计算机器中可调用的 CPU 的个数，其中，每个 CPU 可以分别调用一个线程。

具体地，启动的计算机器中 N 个线程中的每个线程分别得到该线程对应的权重向量切分块与对应的 X 列数据切分块，例如线程 1 对应 X1/W1，线程 2 对应 X2/W2，而线程 3 对应 X3/W3。进而，每个线程可以进行并经计算，根据上述中记载的梯度的计算公式分别计算出对应的逻辑回归梯度 g_1, g_2和g_3。

S105，将 N 个子逻辑回归梯度进行拼接以形成全量逻辑回归梯度。

具体地，将计算得到的逻辑回归梯度g_1, g_2和g_3进行合并，即对逻辑回归梯度g_1, g_2和g_3进行求和，从而可以得到该计算机器的全量逻辑回归梯度g。

本申请实施例的计算方法，将训练数据和权重向量切分成内存中可以存放下的多个小区域，并启动对应个数的线程对个小区域进行并行计算，由此使得计算机器可以支持超大规模逻辑回归模型的训练，提高计算速度，减少训练时间，并且极大地降低计算机器的内存使用量。

图 3 是本申请另一个实施例的逻辑回归梯度的计算方法的流程图。

如图 3 所示，计算方法包括:

S301，获取训练数据，其中，训练数据包括 X 行用户数据和 Y 行点击数据，其中，Y 行点击数据与 X 行用户数据对应。

例如，以以下格式先输入用户数据：

X:[1.3.5.8.9.10.23.22

2.4.6.8.9.12.34.56

……

1.3.6.8.10.11.12]，

其中，第一行为第一个用户数据 X，其中 1.3.5.8.9.10.23.22 代表第一个用户数据 X 中的不同用户特征；第二行为第二个用户数据 X，其中 2.4.6.8.9.12.34.56 代表第二个用户数据 X 中的不同用户特征；而第 N 行为第 N 个用户数据，其中 1.3.6.8.10.11.12 代表第 N 个用户数据中的不同用户特征。

进而，再以以下格式输入点击数据：

Y:[1.0.0.0

2.0.0.0

……

0.0.1.0]，

其中，第一行为第一个点击数据 Y；第二行为第二个点击数据 Y；而第 N 行为第 N 个点击数据。

应当理解的是，上述第一行的第一个用户数据 X 和第一行的第一个点击数据 Y 合并在一起，才是第一个完整的训练数据；第二行的第二个用户数据 X 和第二行的第二个点击数据 Y 合并在一起，才是第二个完整的训练数据；而第 N 行的第 N 个用户数据 X 和第 N 行的第 N 个点击数据 Y 合并在一起，才是第 N 个完整的训练数据。

S302，将 X 行用户数据转换为 X 列数据。

具体地，通过列数据转换装置将 X 行用户数据转换为 X 列数据。

S303，对 X 列数据和权重向量进行切分以形成 N 个 X 列数据切分块和权重向量切分块。

具体地，通过列切分装置将 X 列数据进行切分以得到对应的 N 个 X 列数据切分块。例如，N=3，即将 X 列数据切分为 3 个 X 列数据切分块，这三个 X 列数据切分块分别是:

X1: [1.3.5	X2: [8.9	X3: [10.23.22
2.4.6	8.9	12.34.56
……	……	……
1.3.6]	8]	10.11.12]

进而，如图 2 所示，通过列切分装置将权重向量 W 进行切分以得到对应的权重向量切分块。例如，W1 对应 W 中下标为 1-7 的元素，W2 对应 W 中下标为 8-9 的元素，而 W3 对应 W 中下表为 10-56 的元素。

在本发明的一个实施例中，X 行用户数据中的列数与权重向量的向量 W 宽度相同。也就是说，X 行用户数据的列数和权重向量 W 一样宽，例如，如果权重向量 W 的长度是 100，那么 X 行用户数据中每行用户数据中最大出现的元素也为 100，因此 X 行用户数据中的每行用户数据都是稀疏的，也就是说每个用户数据只包括一部分的特征。例如，湖南的男性用户跟北京的女性用户，前者出现了湖南和男性的特征，后者出现了北京和女性的特征，对应到 X 行用户数据中的两行用户数据中，第一行可能是 1.7，而第二行可能是 9.10。X 行用户数据中 Y 也有两列，第一列 Y 可能是点击，第二列 Y 可能是未点击。

S304，分别启动 N 个线程根据 N 个 X 列数据切分块和权重向量切分块以及对应的 Y 行点击数据生成 N 个子逻辑回归梯度。

在本发明的一个实施例中， N 为计算机器中可调用的 CPU 的个数，其中，每个 CPU 可以分别调用一个线程。

具体地，启动的计算机器中 N 个线程中的每个线程分别得到该线程对应的权重向量切分块与对应的 X 列数据切分块，例如线程 1 对应 X1/W1，

线程 2 对应 X2/W2，而线程 3 对应 X3/W3。进而，每个线程可以进行并经计算，根据上述中记载的梯度的计算公式分别计算出对应的逻辑回归梯度g_1, g_2和g_3。

S305，将 N 个子逻辑回归梯度进行拼接以形成全量逻辑回归梯度。

具体地，将计算得到的逻辑回归梯度g_1, g_2和g_3进行合并，即对逻辑回归梯度g_1, g_2和g_3进行求和，从而可以得到该计算机器的全量逻辑回归梯度g。

S306，将每个机器计算的全量逻辑回归梯度进行汇总以生成最终的全量逻辑回归梯度，其中，计算机器为多个。

具体地，当计算全量逻辑回归梯度的计算机器的个数为个时，可以汇集所有的计算机器计算得到的全量逻辑回归梯度，并将所有的计算机器计算得到的全量逻辑回归梯度进行合并，即将所有的计算机器计算得到的全量逻辑回归梯度加在一起。

S307，根据最终的全量逻辑回归梯度对权重向量进行更新。

具体地，根据以下公式计算权重向量:

第 T+1 轮的权重向量 =第 T 轮的权重向量 W+步长*全量逻辑回归梯度g，

其中，步长为一个浮点数，例如 0.01。

本申请实施例的计算方法，将训练数据和权重向量切分成内存中可以存放下的多个小区域，并启动对应个数的线程对个小区域进行并行计算，进而将每个计算机器计算得到的全量逻辑回归梯度进行汇集，由此可以进一步提高计算速度，减少训练时间。

为了实现上述实施例，本申请还提出一种逻辑回归梯度的计算装置。

图 4 是本申请一个实施例的逻辑回归梯度的计算装置的结构示意图，如图 4 所示，计算装置包括: 获取模块 100、转换模块 200、切分模块 300、生成模块 400 和拼接模块 500。

其中，获取模块 100 用于获取训练数据，其中，训练数据包括 X 行用

户数据和 Y 行点击数据，其中，Y 行点击数据与 X 行用户数据对应。

转换模块 200 用于将 X 行用户数据转换为 X 列数据。

切分模块 300 用于对 X 列数据和权重向量进行切分以形成 N 个 X 列数据切分块和权重向量切分块。其中，N 为计算机器中可调用的 CPU 的个数，X 行用户数据中的列数与权重向量的向量宽度相同。

生成模块 400 用于分别启动 N 个线程根据 N 个 X 列数据切分块和权重向量切分块以及对应的 Y 行点击数据生成 N 个子逻辑回归梯度。

拼接模块 500 用于将 N 个子逻辑回归梯度进行拼接以形成全量逻辑回归梯度。

需要说明的是，前述对逻辑回归梯度的计算方法的解释说明也适用于该实施例的逻辑回归梯度的计算装置，其实现原理类似，此处不再赘述。

本申请实施例的计算装置，将训练数据和权重向量切分成内存中可以存放下的多个小区域，并启动对应个数的线程对个小区域进行并行计算，由此使得计算机器可以支持超大规模逻辑回归模型的训练，提高计算速度，减少训练时间，并且极大地降低计算机器的内存使用量。

图 5 是本申请另一个实施例的逻辑回归梯度的计算装置的结构示意图，如图 5 所示，计算装置包括：获取模块 100、转换模块 200、切分模块 300、生成模块 400、拼接模块 500、汇总模块 600 和更新模块 700。

其中，汇总模块 600 用于将每个机器计算的全量逻辑回归梯度进行汇总以生成最终的全量逻辑回归梯度。

更新模块 700 用于根据最终的全量逻辑回归梯度对权重向量进行更新 。

需要说明的是，前述对逻辑回归梯度的计算方法的解释说明也适用于该实施例的逻辑回归梯度的计算装置，其实现原理类似，此处不再赘述。

本申请实施例的计算装置，将训练数据和权重向量切分成内存中可以存放下的多个小区域，并启动对应个数的线程对个小区域进行并行计算，进而将每个计算机器计算得到的全量逻辑回归梯度进行汇集，由此可以进

一步提高计算速度，减少训练时间。

应当理解，本申请的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中，多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如，如果用硬件来实现，和在另一实施方式中一样，可用本领域公知的下列技术中的任一项或他们的组合来实现：具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路，具有合适的组合逻辑门电路的专用集成电路，可编程门阵列（PGA），现场可编程门阵列（FPGA）等。

在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本申请的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不必须针对的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外，在不相互矛盾的情况下，本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

尽管上面已经示出和描述了本申请的实施例，可以理解的是，上述实施例是示例性的，不能理解为对本申请的限制，本领域的普通技术人员在本申请的范围内可以对上述实施例进行变化、修改、替换和变型。

权 利 要 求 书

1、一种逻辑回归梯度的计算方法，其特征在于，包括以下步骤:

获取训练数据，其中，所述训练数据包括 X 行用户数据和 Y 行点击数据，其中，所述 Y 行点击数据与所述 X 行用户数据对应;

将所述 X 行用户数据转换为 X 列数据;

对所述 X 列数据和权重向量进行切分以形成 N 个 X 列数据切分块和权重向量切分块;

分别启动 N 个线程根据所述 N 个 X 列数据切分块和权重向量切分块以及对应的所述 Y 行点击数据生成 N 个子逻辑回归梯度; 以及

将所述 N 个子逻辑回归梯度进行拼接以形成全量逻辑回归梯度。

2、如权利要求 1 所述的逻辑回归梯度的计算方法，其特征在于，所述 N 为计算机器中可调用的 CPU 的个数。

3、如权利要求 2 所述的逻辑回归梯度的计算方法，其特征在于，所述计算机器为多个，所述方法还包括:

将每个机器计算的全量逻辑回归梯度进行汇总以生成最终的全量逻辑回归梯度。

4、如权利要求 3 所述的逻辑回归梯度的计算方法，其特征在于，还包括:

根据最终的全量逻辑回归梯度对所述权重向量进行更新 。

5、如权利要求 1 所述的逻辑回归梯度的计算方法，其特征在于，所述 X 行用户数据中的列数与所述权重向量的向量宽度相同。

6、一种逻辑回归梯度的计算装置，其特征在于，包括:

获取模块，用于获取训练数据，其中，所述训练数据包括 X 行用户数据和 Y 行点击数据，其中，所述 Y 行点击数据与所述 X 行用户数据对应;

转换模块，用于将所述 X 行用户数据转换为 X 列数据;

切分模块，用于对所述 X 列数据和权重向量进行切分以形成 N 个 X 列数据切分块和权重向量切分块;

生成模块，用于分别启动 N 个线程根据所述 N 个 X 列数据切分块和权重向量切分块以及对应的所述 Y 行点击数据生成 N 个子逻辑回归梯度;以及

拼接模块，用于将所述 N 个子逻辑回归梯度进行拼接以形成全量逻辑回归梯度。

7、如权利要求 6 所述的逻辑回归梯度的计算装置，其特征在于，所述 N 为计算机器中可调用的 CPU 的个数。

8、如权利要求 7 所述的逻辑回归梯度的计算装置，其特征在于，所述计算机器为多个，还包括:

汇总模块，用于将每个机器计算的全量逻辑回归梯度进行汇总以生成最终的全量逻辑回归梯度。

9、如权利要求 8 所述的逻辑回归梯度的计算装置，其特征在于，还包括:

更新模块，用于根据最终的全量逻辑回归梯度对所述权重向量进行更新。

10、如权利要求 6 所述的逻辑回归梯度的计算装置，其特征在于，所述 X 行用户数据中的列数与所述权重向量的向量宽度相同。

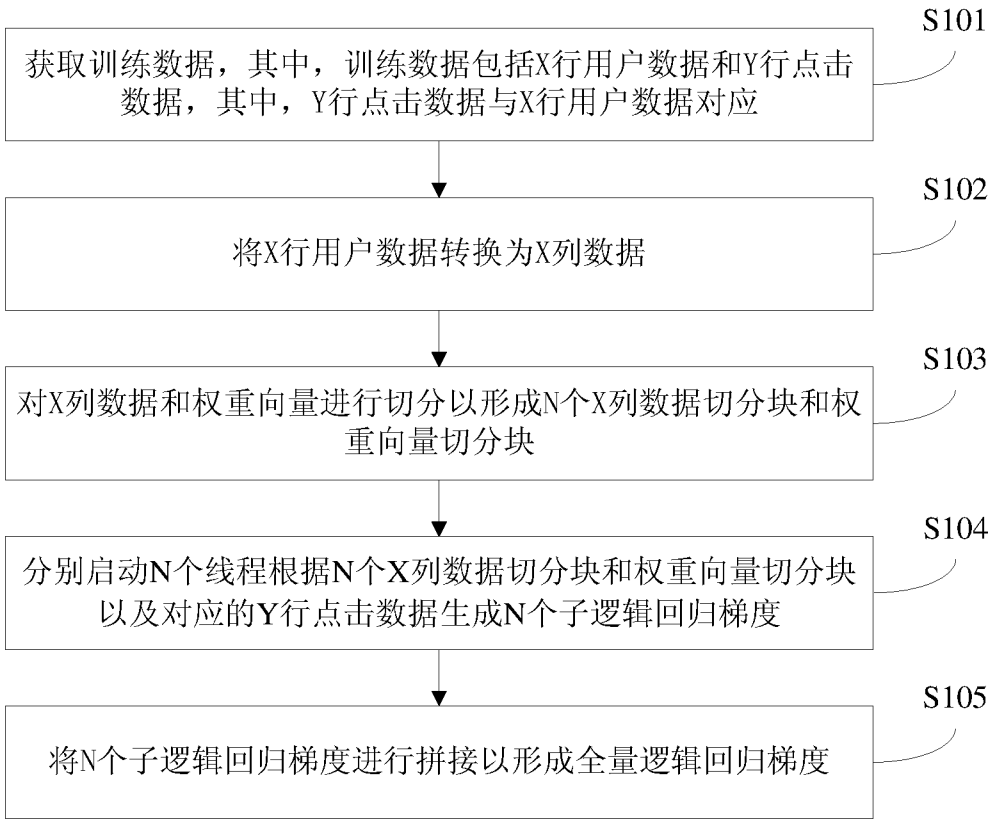

S101
获取训练数据，其中，训练数据包括X行用户数据和Y行点击数据，其中，Y行点击数据与X行用户数据对应
S102
将X行用户数据转换为X列数据
S103
对X列数据和权重向量进行切分以形成N个X列数据切分块和权重向量切分块
S104
分别启动N个线程根据N个X列数据切分块和权重向量切分块以及对应的Y行点击数据生成N个子逻辑回归梯度
S105
将N个子逻辑回归梯度进行拼接以形成全量逻辑回归梯度

图 1

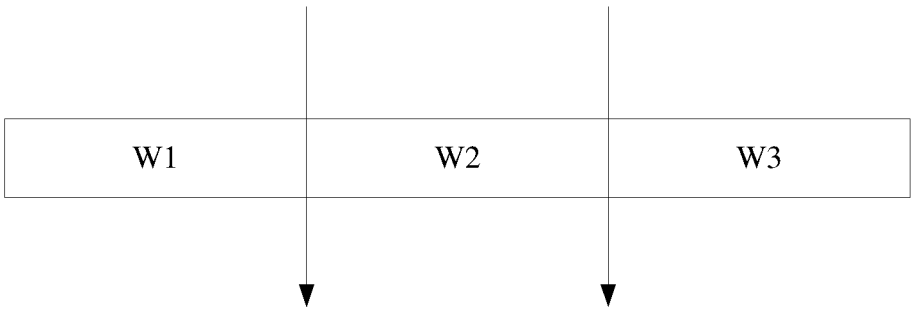

W1
W2
W3

图 2

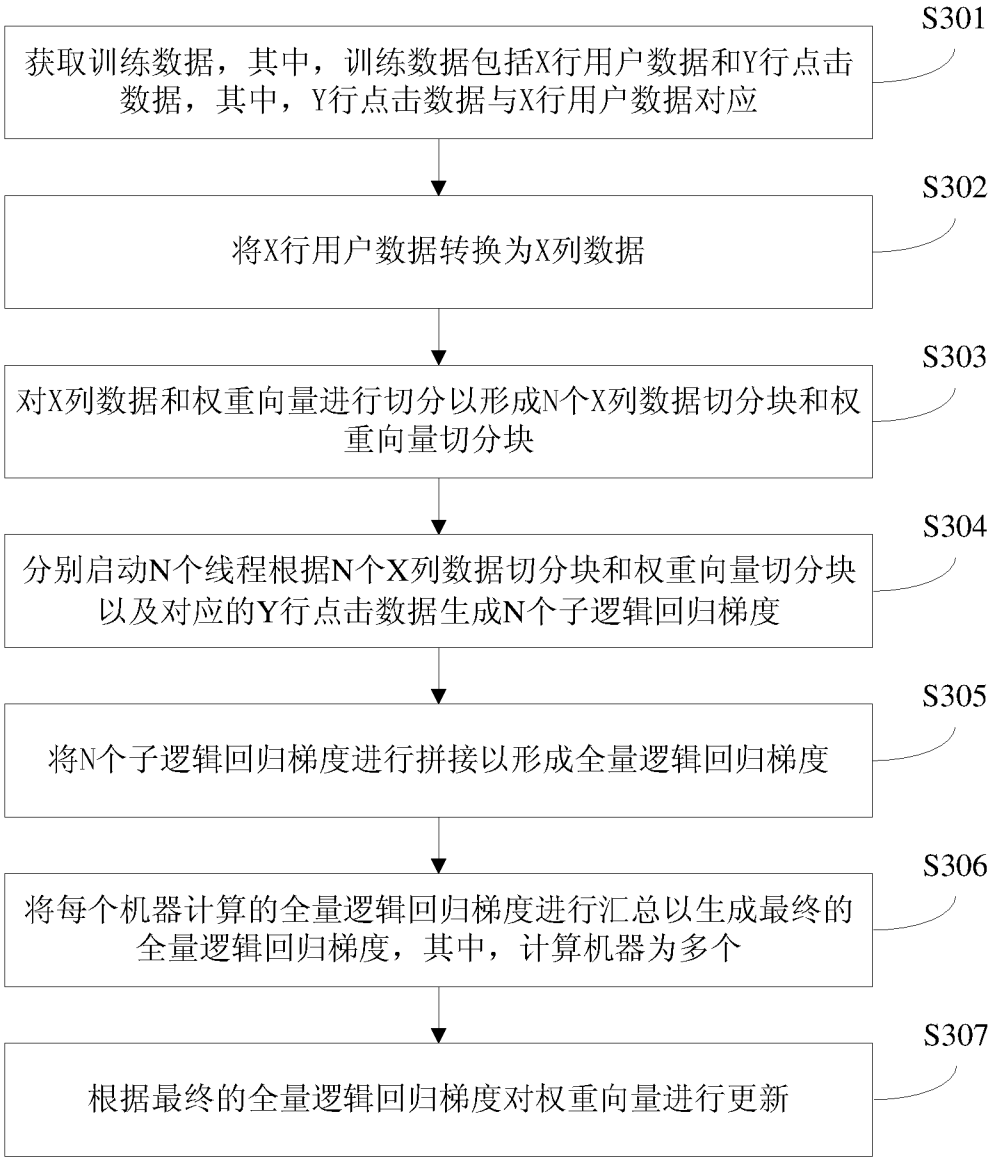

S301
获取训练数据，其中，训练数据包括X行用户数据和Y行点击数据，其中，Y行点击数据与X行用户数据对应
S302
将X行用户数据转换为X列数据
S303
对X列数据和权重向量进行切分以形成N个X列数据切分块和权重向量切分块
S304
分别启动N个线程根据N个X列数据切分块和权重向量切分块以及对应的Y行点击数据生成N个子逻辑回归梯度
S305
将N个子逻辑回归梯度进行拼接以形成全量逻辑回归梯度
S306
将每个机器计算的全量逻辑回归梯度进行汇总以生成最终的全量逻辑回归梯度，其中，计算机器为多个
S307
根据最终的全量逻辑回归梯度对权重向量进行更新

图 3

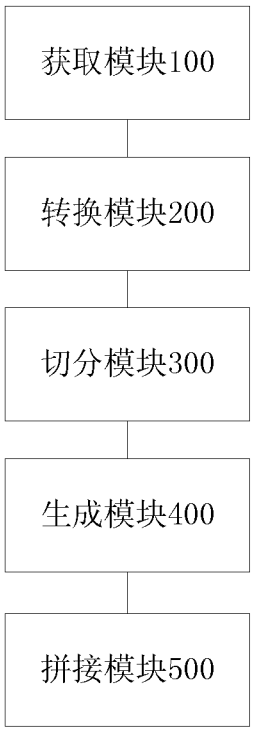

获取模块100
转换模块200
切分模块300
生成模块400
拼接模块500

图 4

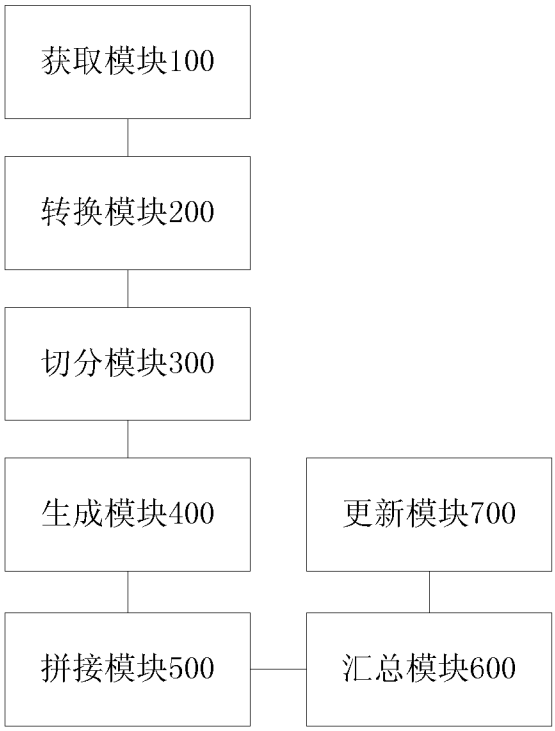

获取模块100
转换模块200
切分模块300
生成模块400
更新模块700
拼接模块500
汇总模块600

图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/071130

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/18 (2006.01) 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)

G06F

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: logistic w regression, gradient, weight+, training, sample, parallel. , division, block

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 104536983 A (BEIJING ZHANGKUO TECHNOLOGY CO., LTD.) 22 April 2015 (22.04.2015) description, paragraphs [0050]-[0077], and figures 1-3	1-10
A	CN 103745225 A (BEIJING GEO POLYMERIZATION NETWORK TECHNOLOGY CO., LTD.) 23 April 2014 (23.04.2014) the whole document	1-10
A	CN 105045819 A (TENCENT INCORPORATION) 11 November 2015 (11.11.2015) the whole document	1-10
A	US 2009248595 A1 (LU, YUMAO et al.) 01 October 2009 (01.10.2009) the whole 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	Date of mailing of the international search report
07 March 2017	30 March 2017
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer JIN, Xia Telephone No. (86-10) 62414438

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2017/071130

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 104536983 A	22 April 2015	None	
CN 103745225 A	23 April 2014	None	
CN 105045819 A	11 November 2015	None	
US 2009248595 A1	01 October 2009	None	

国际检索报告

国际申请号
PCT/CN2017/071130

A. 主题的分类

G06F 17/18(2006.01)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)

G06F

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;EPODOC;CNKI;IEEE: 逻辑回归，梯度，权重，训练，并行，分块，切分，logistic w regression, gradient，weight+，training，sample，parallel+，division，block

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
A	CN 104536983 A (北京掌阔技术有限公司) 2015年 4月 22日 (2015-04-22) 说明书第[0050]-[0077]段，图1-3	1-10
A	CN 103745225 A (北京集奥聚合网络技术有限公司) 2014年 4月 23日 (2014-04-23) 全文	1-10
A	CN 105045819 A (深圳市腾讯计算机系统有限公司) 2015年 11月 11日 (2015-11-11) 全文	1-10
A	US 2009248595 A1 (LU, YUMAO 等) 2009年 10月 1日 (2009-10-01) 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	国际检索报告邮寄日期
2017年 3月 7日	2017年 3月 30日

ISA/CN的名称和邮寄地址	受权官员
中华人民共和国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	金霞
传真号 (86-10)62019451	电话号码 (86-10)62414438

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2017/071130

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 104536983 A	2015年 4月 22日	无	
CN 103745225 A	2014年 4月 23日	无	
CN 105045819 A	2015年 11月 11日	无	
US 2009248595 A1	2009年 10月 1日	无	