

(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
2016 年 7 月 14 日 (14.07.2016)



(10) 国际公布号
WO 2016/110125 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2015/089607
- (22) 国际申请日: 2015 年 9 月 15 日 (15.09.2015)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510012075.6 2015 年 1 月 9 日 (09.01.2015) CN
201510013258.X 2015 年 1 月 9 日 (09.01.2015) CN
- (71) 申请人: 北京大学 (PEKING UNIVERSITY)
[CN/CN]; 中国北京市海淀区颐和园路 5 号, 北京大学, Beijing 100871 (CN)。
- (72) 发明人: 段凌宇 (DUAN, Lingyu); 中国北京市海淀区颐和园路 5 号, 北京大学, Beijing 100871 (CN)。
王哲 (WANG, Zhe); 中国北京市海淀区颐和园路 5 号, 北京大学, Beijing 100871 (CN)。
王潇放 (WANG, Xiaofang); 中国北京市海淀区颐和园路 5 号, 北京大学, Beijing 100871 (CN)。
黄铁军 (HUANG, Tiejun); 中国北京市海淀区颐和园路 5 号, 北京大学, Beijing 100871 (CN)。
高文 (GAO, Wen);

中国北京市海淀区颐和园路 5 号, 北京大学, Beijing 100871 (CN)。

(74) 代理人: 北京路浩知识产权代理有限公司 (CN-KNOWHOW INTELLECTUAL PROPERTY AGENT LIMITED); 中国北京市海淀区丹棱街 3 号中国电子大厦 B 座 18 层, Beijing 100080 (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, 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, KN, KP, KR, 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: HASH METHOD FOR HIGH DIMENSION VECTOR, AND VECTOR QUANTIZATION METHOD AND DEVICE

(54) 发明名称: 高维向量的哈希方法、向量量化方法及装置

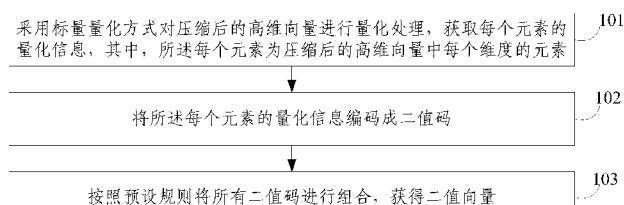


图 1

- 101 Perform quantization processing on a compressed high dimension vector in a scalar quantization manner, to acquire quantization information of each element, each element being an element of each dimension in the compressed high dimension vector
- 102 Code the quantization information of each element into binary code
- 103 Combine all binary code according to a preset rule, to obtain a binary vector

(57) Abstract: A hash method for a high dimension vector, and a vector quantization method and device. The hash method comprises: performing quantization processing on a compressed high dimension vector in a scalar quantization manner, to acquire quantization information of each element, each element being an element of each dimension in the compressed high dimension vector (S101); coding the quantization information of each element into binary code (S102); and combing all binary code according to a preset rule, to obtain a binary vector (S103). The method can reduce storage complexity in a neighbor query process in the prior art, and improve search accuracy and search efficiency.

(57) 摘要: 一种高维向量的哈希方法、向量量化方法及装置, 所述哈希方法包括: 采用标量量化方式对压缩后的高维向量进行量化处理, 获取每个元素的量化信息, 其中, 所述每个元素为压缩后的高维向量中每个维度的元素(S101); 将所述每个元素的量化信息编码成二进制码(S102); 按照预设规则将所有二进制码

进行组合, 获得二值向量(S103)。上述方法能够对降低现有技术中近邻查询过程中的存储复杂度, 以及提升检索准确率和检索效率。



WO 2016/110125 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))。

高维向量的哈希方法、向量量化方法及装置

技术领域

5 本发明涉及计算机技术，具体涉及一种高维向量的哈希方法、向量量化方法及装置。

背景技术

10 当前，高维向量的近邻查询在计算机的图像搜索中提供着关键性的技术支撑。高维向量的近邻查询是指：给定一个高维向量，从一个向量集中找出与其最近的若干个向量。高维向量的近邻查询可应用于人工智能、机器学习以及计算机视觉等方面。

为此，越来越多的业内人士研究高维向量的近邻查询的相关技术。现有技术中，解决近邻查询的方法主要分为两类。第一类方法是通过建立索引结构，离线将相似的向量集中到一起，在线查询时，从索引中快速的读取结果，从而获得查询向量的近邻。然而，该第一类方法最大的缺点是内存消耗大，
15 该索引结构的内存占用量比原始数据的内存占用量还大，导致查询效率非常低。

第二类方法是通过乘积量化的方式将高维向量分成多段，每段单独量化，并建立索引。以一个维度为 8 的向量为例 $Y=\{y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8\}$ 。常用的乘积量化会将这个 8 维向量划分成两段，分别是 $\{y_1, y_2, y_3, y_4\}$ 以及 $\{y_5, y_6, y_7, y_8\}$ 。然后每段的维度统一共同量化。该方法的缺点是需要
20 将高维向量拆分成许多段，每段使用额外的查找表；当向量维度较高时，会破坏向量的空间结构，内存占用大，导致采用量化后的信息在图像检索中检索准确率降低，且检索速度慢，效率低。

为此，如何降低近邻查询过程中的存储复杂度，并提升检索准确率和检索效率成为当前需要解决的技术问题。
25

发明内容

针对现有技术中的缺陷，本发明提供了一种高维向量的哈希方法及装置，能够降低现有技术中近邻查询过程中的存储复杂度，以及提升检索准确率和检索效率。

第一方面，本发明提供一种高维向量的哈希方法，包括：

采用标量量化方式对压缩后的高维向量进行量化处理，获取每个元素的量化信息，其中，所述每个元素为压缩后的高维向量中每个维度的元素；

将所述每个元素的量化信息编码成二值码；

5 按照预设规则将所有二值码进行组合，获得二值向量。

可选地，采用标量量化方式对压缩后的高维向量进行量化处理之前，所述方法还包括：

对待处理的高维向量进行压缩处理，获取压缩后的高维向量。

10 可选地，采用标量量化方式对压缩后的高维向量进行量化处理，获取每个元素的量化信息，包括：

获取压缩后的高维向量中每个维度的元素；

根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得每个元素的量化信息；

所述每个维度对应的区间是根据预设的训练数据集训练获取的。

15 可选地，根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间之前，所述方法还包括：

获取所述压缩后高维向量的维度，根据所述维度获取与所述维度对应的区间信息，所述区间信息包括：预先获取的所述维度中每个维度对应的区间。

20 可选地，根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间之前，所述方法还包括：

根据预设的训练数据集训练获取每个维度对应的区间，所述每个维度为所述高维向量中的每个维度；

每一维度对应的区间均是不重叠的。

可选地，根据预设的训练数据集训练获取每个维度对应的区间，包括：

25 获取训练数据集中所有向量的每一维度的元素集合，所述元素集合包括该维度上所有向量的元素，所述训练数据集中所有向量的维度相同；

根据预设距离差和量化误差，获取用于划分区间的目标函数；

根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间。

30 可选地，根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间，包括：

枚举所述元素集合中区间划分的边界值，将每组所述区间划分的边界值

作为所述目标函数的参数值，确定并记录所述目标函数对应的取值；

选择使目标函数取得极值的边界值划分该维度，获得该维度对应的区间。

可选地，给每个维度对应的区间的元素进行编码，被量化到某个区间的元素将编码成该区间对应的二值码。编码的方式可以采用自然二进制码和一
5 元码。

可选地，按照预设规则将所有二值码进行组合，获得二值向量，包括：

将所有二值码按照每一二值码对应的维度序号顺序拼接，获得二值向量。

可选地，所述对待处理的高维向量进行压缩处理，获取压缩后的高维向量，包括：

10 采用预设的线性变换矩阵对待处理的高维向量进行压缩处理，获得压缩后的高维向量；

所述线性变换矩阵为通过训练样本预先训练获取的。

可选地，所述对待处理的高维向量进行压缩处理，获取压缩后的高维向量，包括：

15 采用预设的非线性变换函数对待处理的高维向量进行压缩处理，获得压缩后的高维向量；

所述非线性变换函数通过训练样本预先训练获取的。

第二方面，本发明还提供一种高维向量的近邻查询方法，包括：

采用如上任一所述的方法获取待处理的高维向量的二值向量；

20 获取所述二值向量的哈希二值码；

根据所述哈希二值码的编码方式，选择与该编码方式对应的度量距离，采用所述度量距离度量所述二值向量与其他高维向量对应的二值向量的相似度。利用二值向量的相似性/相似度表达高维向量之间的相似性。

第三方面，本发明还提供一种高维向量的哈希装置，包括：

25 量化单元，用于采用标量量化方式对压缩后的高维向量进行量化处理，获取每个元素的量化信息，其中，所述每个元素为压缩后的高维向量中每个维度的元素；

编码单元，用于将所述每个元素的量化信息编码成二值码；

组合单元，用于按照预设规则将所有二值码进行组合，获得二值向量。

30 可选地，所述装置还包括：

压缩单元，用于在量化单元量化之前对待处理的高维向量进行压缩处理，获取压缩后的高维向量。

可选地，量化单元，具体用于

获取压缩后的高维向量中每个维度的元素；

根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得每个元素的量化信息；

5 所述每个维度对应的区间是根据预设的训练数据集训练获取的。

可选地，量化单元，具体用于

获取压缩后的高维向量中每个维度的元素；

获取所述压缩后高维向量的维度，根据所述维度，获取与所述维度对应的区间信息，所述区间信息包括：预先获取的所述维度中每个维度对应的区
10 间；

根据获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得每个元素的量化信息；

所述每个维度对应的区间是根据预设的训练数据集训练获取的。

可选地，所述量化单元，还用于

15 根据预设的训练数据集训练获取每个维度对应的区间，所述每个维度为所述高维向量中的每个维度；

每一维度对应的区间均是不重叠的。

可选地，所述量化单元，具体用于

20 获取训练数据集中所有向量的每一维度的元素集合，所述元素集合包括该维度上所有向量的元素，所述训练数据集中所有向量的维度相同；

根据预设距离差和量化误差，获取用于划分区间的目标函数；

根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间。

可选地，所述量化单元，具体用于

25 枚举所述元素集合中区间划分的边界值，将每组所述区间划分的边界值作为所述目标函数的参数值，确定并记录所述目标函数对应的取值；

选择使目标函数取得极值的边界值划分该维度，获得该维度对应的区间。

可选地，组合单元，具体用于

将所有二值码按照每一二值码对应的维度序号顺序拼接，获得二值向量。

30 可选地，所述编码单元，具体用于

对每一维度的区间（即量化后的区间）中的元素编码成二值码，编码的方式可以是自然二进制码也可以是一元码。

可选地，压缩单元，具体用于

采用预设的线性变换矩阵对待处理的高维向量进行压缩处理，获得压缩后的高维向量；

所述线性变换矩阵为通过训练样本预先训练获取的。

5 可选地，压缩单元，具体用于

采用预设的非线性变换函数对待处理的高维向量进行压缩处理，获得压缩后的高维向量；

所述非线性变换函数为通过训练样本预先训练获取的。

第四方面，本发明提供一种高维向量的近邻查询装置，包括：

10 二值向量获取单元，用于采用上述任一所述的装置获取待处理的高维向量的二值向量；

哈希二值码获取单元，用于获取所述二值向量的哈希二值码；

15 相似度度量单元，用于根据所述哈希二值码的编码方式，选择与该编码方式对应的度量距离，采用所述度量距离度量所述二值向量与其他高维向量对应的二值向量的相似度。

第五方面，本发明提供一种向量量化方法，包括：

获取待量化的向量中每个维度的元素；

根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得所述待量化的向量的量化信息；

20 所述每个维度对应的区间是根据预设的训练数据集训练获取的。

可选地，所述根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间之前，所述方法还包括：

获取所述待量化向量的维度，根据所述维度获取与所述维度对应的区间信息，所述区间信息包括：预先获取的所述维度中每个维度对应的区间。

25 可选地，所述根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间之前，所述方法还包括：

根据预设的训练数据集训练获取每个维度对应的区间，所述每个维度为所述待量化向量中的每个维度；

每一维度对应的区间均是不重叠的。

30 可选地，根据预设的训练数据集训练获取每个维度对应的区间，包括：

获取训练数据集中所有向量的每一维度的元素集合，所述元素集合包括该维度上所有向量的元素，所述训练数据集中所有向量的维度相同，且与待

量化向量的维度相同；

根据预设距离差和量化误差，获取用于划分区间的目标函数；

根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间。

- 5 可选地，根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间，包括：

枚举所述元素集合中区间划分的边界值，将每组所述区间划分的边界值作为所述目标函数的参数值，确定并记录所述目标函数对应的取值；

选择使目标函数取得极值的边界值划分该维度，获得该维度对应的区间。

- 10 第六方面，本发明还提供一种向量量化装置，包括：

元素获取单元，用于获取待量化的向量中每个维度的元素；

量化单元，用于根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得所述待量化的向量的量化信息；

其中，所述每个维度对应的区间是根据预设的训练数据集训练获取的。

- 15 可选地，所述装置还包括：

维度确定单元，用于获取所述待量化向量的维度，根据所述维度，获取与所述维度对应的区间信息，所述区间信息包括：预先获取的所述维度中每个维度对应的区间。

可选地，所述装置还包括：

- 20 维度划分单元，用于根据预设的训练数据集训练获取每个维度对应的区间，所述每个维度为所述待量化向量中的每个维度；

每一维度对应的区间均是不重叠的。

可选地，维度划分单元，具体用于

- 25 获取训练数据集中所有向量的每一维度的元素集合，所述元素集合包括该维度上所有向量的元素，所述训练数据集中所有向量的维度相同，且与待量化向量的维度相同；

根据预设距离差和量化误差，获取用于划分区间的目标函数；

根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间。

- 30 可选地，维度划分单元，具体用于

获取训练数据集中所有向量的每一维度的元素集合，所述元素集合包括该维度上所有向量的元素，所述训练数据集中所有向量的维度相同，且与待

量化向量的维度相同；

根据预设距离差和量化误差，获取用于划分区间的目标函数；

枚举所述元素集合中区间划分的边界值，将每组所述区间划分的边界值作为所述目标函数的参数值，确定并记录所述目标函数对应的取值；

5 选择使目标函数取得极值的边界值划分该维度，获得该维度对应的区间。

由上述技术方案可知，本发明的高维向量的哈希方法及装置，通过获取高维向量中每个元素的量化信息，进而将量化信息编码成二值码，由此，可组合二值码，获得高维向量的二值向量，进而能够对降低现有技术中近邻查询过程中的存储复杂度，以及提升检索准确率和检索效率。

10 本发明的向量量化方法及装置，通过获取待量化的向量中每个维度的元素，进而将元素量化到该元素所属维度对应的区间，获得待量化的向量的量化信息，由此，可实现对高维向量进行量化处理，且采用量化后的信息能够提高图像检索中的检索准确率。

附图说明

15 图1为本发明一实施例提供的高维向量的哈希方法的流程示意图；
图2为本发明一实施例提供的高维向量的哈希方法的流程示意图；
图3为本发明一实施例提供的高维向量的近邻查询方法的流程示意图；
图4为本发明一实施例提供的高维向量的哈希装置的结构示意图；
图5为本发明一实施例提供的高维向量的近邻查询装置的结构示意图；
20 图6为本发明一实施例提供的向量量化方法的流程示意图；
图7为本发明另一实施例提供的向量量化方法的流程示意图；
图8为本发明一实施例提供的向量量化装置的结构示意图。

具体实施方式

下面结合附图，对发明的具体实施方式作进一步描述。以下实施例仅用于更加清楚地说明本发明的技术方案，而不能以此来限制本发明的保护范围。本发明实施例中所使用的“第一”、“第二”仅为更清楚的说明本申请的内容，不具有特定含义，也不限定任何内容。

当前，业内还出现一种近邻查询的技术方案，例如，该技术方案是通过压缩高维向量降低高维向量的维度然后再将降维后的向量量化成二值码的哈希方法。不同于现有技术中提及的第一类和第二类方法，该技术方案
30 中的哈希方法通常不需要额外的索引，直接利用量化后的二值码之间的距

离判定向量之间的相似性。

然而，该近邻查询的技术方案还仅仅是简单地将压缩后向量的每个元素用一个预设的阈值量化成单个比特，这个过程会引入较大的量化误差。

为此，如何降低近邻查询过程中的存储复杂度，提升检索效率；以及解决上述方案中单个比特量化的误差问题，本发明实施例中引入更精确的量化的机制（比如多比特量化）进而可减少量化误差。

图1示出了本发明一实施例提供的高维向量的哈希方法的流程示意图，如图1所示，本实施例中的高维向量的哈希方法如下所述。

101、采用标量量化方式对压缩后的高维向量进行量化处理，获取每个元素的量化信息，其中，所述每个元素为压缩后的高维向量中每个维度的元素。

举例来说，对于压缩后的高维向量 $X=[x_1, x_2, \dots, x_m]$ 中的任何一个元素 x_i ($1 \leq i \leq m$)， x_i 量化后的量化信息为 $Q(x_i)$ ，其中， Q 表示量化函数。

102、将所述每个元素的量化信息编码成二值码；

在具体应用中，每个维度的元素量化后的区间编码即为该元素的编码。

15 对于元素 x_i ，若 $Q(x_i)=S_{ij}$ ，那么 x_i 将被编码成 $B(S_{ij})$ 。

举例来说，可在应用中，将每个维度的元素 x_i 划分到区间，将每个维度的元素划分后的区间 S_{ij} ， $1 \leq i \leq m$ ， $1 \leq j \leq u$ ，编码成二值码 $B(S_{ij})$ 。

其中， m 为压缩后的高维向量的维度， u 为非零自然数， u 表示的是每个维度划分的区间数，比如 $u=4$ ，那么就是将每个维度划分成4个区间，此时若采用二进制编码，可以分别将每个区间段依次编码成00，01，10，11等。

具体的，二值码包括但不限于自然二进制码，包括：

举例来说，设 $u=4$ ，如果将 S_{ij} 编码成2个比特，那么 S_{i1} ， S_{i2} ， S_{i3} ， S_{i4} 对应的自然二进制码为00，01，10，11。对于二进制编码，采用的距离度量为二值码对应数值的差。比如00与01的距离 $|00_2 - 01_2| = |0 - 1| = 1$ 。

25 当然，在实际应用中，二值码包括但不限于上述的二进制码，还可为一元码，例如：设 $u=4$ ，如果将 S_{ij} 编码成3个比特，那么 S_{i1} ， S_{i2} ， S_{i3} ， S_{i4} 对应的一元码为000，001，011，111。对于一元码，采用的距离度量可以为汉明距离。汉明距离的定义如下：给定两个二值码，其汉明距离等于异或以后数值为1的比特位的数目。比如010与011的抑或以后等于001，那么010与011的汉

明距离为1。

其它编码方式在本实施例中也可使用，本实施例仅为举例说明，并不是限定性的内容。

103、按照预设规则将所有二值码进行组合，获得二值向量。

5 例如，可将所有二值码按照每一二值码对应的维度序号顺序拼接，获得二值向量。

通常，每一二值码对应的维度序号即为该二值码的下标，进而可按照所有二值码的下标顺序拼接获得二值向量。举例来说，对于一个向量 (y_1, y_2, y_3) ，维度序号就是下标，该下标实质上就是1, 2, 3这些序号。其中 y_1 量化成
10 00, y_2 量化成01, y_3 量化成11，那么拼接到一起获得二值向量就是000111。为此，可按维度序号从左至右拼接到一起。

首先，按顺序排列每个元素的量化信息编码后的二值码。对于 X ，按下标顺序排列 X 中各个元素的量化信息编码后的二值码 $B(Q(x_1))$, $B(Q(x_2))$, ..., $B(Q(x_m))$ 。

15 接着，把各个元素的量化信息编码后的二值码组合到一起，构成二值向量。例如，将 $B(Q(x_1))$, $B(Q(x_2))$, ..., $B(Q(x_m))$ 组合，得到二值向量 H , $H=[B(Q(x_1)), B(Q(x_2)), \dots, B(Q(x_m))]$ 。

在本实施例中二值向量即为高维向量哈希后的结果。

20 本实施例的高维向量的哈希方法可通过获取高维向量中每个元素的量化信息，进而将量化信息编码成二值码，由此，可组合二值码，获得高维向量的二值向量，进而能够对降低现有技术中近邻查询过程中的存储复杂度，以及提升检索效率。

图2示出了本发明一实施例提供的高维向量的哈希方法的流程示意图，如图2所示，本实施例中的高维向量的哈希方法如下所述。

25 201、对待处理的高维向量进行压缩处理，获取压缩后的高维向量。

例如，采用预设的线性变换矩阵对待处理的高维向量进行压缩处理，获得压缩后的高维向量；

所述线性变换矩阵为通过训练样本预先训练获取的。

举例来说，步骤201的具体实现过程如下：

A01、采集一个离线训练的数据集，组成训练样本 S_Z ，例如，采集一个 n 维向量集合 $S_Z=\{Z_1, Z_2, \dots, Z_t\}$ ， $Z_1, Z_2, \dots, Z_t$ 为 S_Z 中的元素。

其中， Z_i 为 n 维向量， $1 \leq i \leq t$ ， t 为自然数。

5 A02、从训练样本 S_Z 中训练获取一个线性变换矩阵 A 。

即，可采用主要成分分析方法（Principal Component Analysis，简称PCA）从向量集合 S_Z 中学习获得一个线性变换矩阵 A ，该线性变换矩阵 A 的维度为 $n*m$ ， m 为压缩后的高维向量的维度。

A03、将待处理的高维向量乘以线性变换矩阵，得到压缩后的高维向量。

10 例如，可对于待处理的高维向量 $Y=[y_1, y_2, \dots, y_n]$ ，乘以线性变换矩阵 A ，得到压缩后的高维向量 $X=F(Y)=(Y^T A)^T=[x_1, x_2, \dots, x_m]$ ， m 为自然数， F 代表压缩处理。

也可以将待处理的高维向量进行非线性变换，得到压缩后的高维向量。

15 例如可对于待处理的高维向量 $Y=[y_1, y_2, \dots, y_n]$ ，进行非线性函数 $\cos()$ 变换，得到压缩后的高维向量 $X=K(Y)=\cos(YS)=[\cos(y_1 s_1), \cos(y_2 s_2), \dots, \cos(y_m s_m)]$ ， m 为自然数， K 为非线性变换。非线性函数为预先通过训练样本获取的。

当然，在实际应用中，还可采用其他方式对待处理的高维向量进行压缩处理，不限定上述举例的步骤A01至步骤A03。

202、获取压缩后的高维向量中每个维度的元素。

20 203、根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得每个元素的量化信息。

例如，对于 X 中的任何一个元素 x_i （ $1 \leq i \leq m$ ）， x_i 量化后的量化信息为 $Q(x_i)$ ，其中， Q 表示量化函数。

25 在具体实现过程中，预先将每个维度的元素取值范围划分成几个独立的区间。设 X 中每个维度的元素 x_i 取值范围为 $L_i=[l_i, r_i]$ 。将 L_i 划分成 u 个独立的区间， $S_{i1}, S_{i2}, \dots, S_{iu}$ 。

其中，区间之间不重叠。且 $L_i = S_{i1} \cup S_{i2} \cup \dots \cup S_{iu}$ 。

其次，把每个元素量化到该元素所属维度对应的区间。

若元素 x_i 属于 S_{ij} , 那么 x_i 将被量化到 $S_{ij}=Q(x_i)$ 。

其中, 每个维度对应的区间是根据预设的训练数据集训练获取的。

举例来说, 可预先离线获取所述压缩后高维向量的维度, 根据所述维度获取与所述维度对应的区间信息, 所述区间信息包括: 预先获取的所述维度
5 中每个维度对应的区间, 进而再进一步执行步骤203。

例如, 高维向量可为128维, 则可获取与128维对应的每一维度的区间信息。在部分实施例中, 不同维度的区间信息可不同。

204、将所述每个元素的量化信息编码成二值码。

例如, 对于量化信息 $Q(x_i)$ 编码成二值码 $B(Q(x_i))$, B 表示编码函
10 数。

205、按照预设规则将所有二值码进行组合, 获得二值向量。

例如, 将编码后的二值码按照下标顺序拼接 $B(Q(x_1)), B(Q(x_2)), \dots, B(Q(x_m))$, 得到 $H=[B(Q(x_1)), B(Q(x_2)), \dots, B(Q(x_m))]$, H 为二值向量。

15 本实施例中的 H 为紧凑的二值向量。

本实施例的高维向量的哈希方法能够对降低现有技术中近邻查询过程中的存储复杂度, 以及提升检索效率。

本实施例能够有效解决现有哈希方法中的单比特量化的性能损失问题。现有的哈希方法大多是简单地将压缩后向量的每个元素用一个预设的阈值
20 量化成单个比特, 这个过程会引入较大的量化误差。而本实施例则提供了一种更科学并且支持多比特量化的量化机制, 能够大大降低量化误差, 解决现有哈希方法中单比特量化的问题。

下面对上述步骤203进行详细说明如下:

25 B01、根据预设的训练数据集训练获取每个维度对应的区间, 所述每个维度为所述高维向量中的每个维度; 每一维度对应的区间均是不重叠的。

B02、根据预先获取的每个维度对应的区间, 将所述元素量化到该元素所属维度对应的区间。

该步骤B02的具体实现过程如下:

30 B021、获取训练数据集中所有向量的每一维度的元素集合, 所述元素集合包括该维度上所有向量的元素, 所述训练数据集中所有向量的维度相同;

B022、根据预设距离差和量化误差, 获取用于划分区间的目标函数;

B023、根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间。

具体地，B0231、枚举所述元素集合中区间划分的边界值，将每组所述区间划分的边界值作为所述目标函数的参数值，确定并记录所述目标函数对应的取值；

B0232、选择使目标函数取得极值（如最小值）的边界值划分该维度，获得该维度对应的区间。

可理解的是，在具体应用中，目标函数还可能非连续而且不可导的函数，在应用中存在无法直接获取该目标函数的极值，也就是说，无法直接判断该目标函数在某个取值下是否是极大或者极小值。为此，可预先枚举足够的边界值，从中选择使目标函数相对最大（或者最小）的边界值来划分区间。

当然，可根据实际需要选择使目标函数取得最大值或最小值的方式，本实施例均为举例说明，不对其进行限定。

图3示出了本发明一实施例提供的高维向量的近邻查询方法的流程示意图，如图3所示，本实施例中的高维向量的近邻查询方法如下所述。

301、获取待处理的高维向量的二值向量。

在本实施例中可采用上述任一实施例所述的方法获取待处理的高维向量的二值向量。

302、获取所述二值向量的哈希二值码；

303、根据所述哈希二值码的编码方式，选择与该编码方式对应的度量距离，采用所述度量距离度量所述二值向量与其他高维向量对应的二值向量的相似度。

例如，对于两个紧凑的二值向量H1和H2，选择与所用二值码相匹配的距离D，来度量二值向量之间的相似性。

即两个二值向量H1和H2的相似性用D(H1, H2)来表示。

对于两个二值向量H1和H2，其相似性的度量举例可以为汉明距离，也可以为曼哈顿举例度量，其他举例度量方法本实施例不对其进行限定。

上述近邻查询方法可降低现有技术中高额的存储复杂度，并且大幅度地提升检索的效率。

图4示出了本发明一实施例提供的高维向量的哈希装置的结构示意图，如图4所示，本实施例的高维向量的哈希装置包括：量化单元41、编码单元42和组合单元43；

其中，量化单元41用于采用标量量化方式对压缩后的高维向量进行量化处理，获取每个元素的量化信息，其中，所述每个元素为压缩后的高维向量中每个维度的元素；

编码单元42用于将所述每个元素的量化信息编码成二值码；

5 组合单元43用于按照预设规则将所有二值码进行组合，获得二值向量。

在具体应用中，前述的装置还可包括图中未示出的压缩单元40，该压缩单元40用于在量化单元41量化之前对待处理的高维向量进行压缩处理，获取压缩后的高维向量。例如，采用预设的线性变换矩阵对待处理的高维向量进行压缩处理，获得压缩后的高维向量；所述线性变换矩阵为通过训练样本预
10 先训练获取的。

在一种可能的应用中，量化单元41可具体用于，获取压缩后的高维向量中每个维度的元素；

根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得每个元素的量化信息；

15 所述每个维度对应的区间是根据预设的训练数据集训练获取的。

在另一种具体的应用中，量化单元41可具体用于，获取压缩后的高维向量中每个维度的元素；

获取所述压缩后高维向量的维度，根据所述维度，获取与所述维度对应的区间信息，所述区间信息包括：预先获取的所述维度中每个维度对应的区
20 间；

根据获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得每个元素的量化信息；

所述每个维度对应的区间是根据预设的训练数据集训练获取的。

25 举例来说，所述量化单元41还可用于，根据预设的训练数据集训练获取每个维度对应的区间，所述每个维度为所述高维向量中的每个维度；

例如，获取训练数据集中所有向量的每一维度的元素集合，所述元素集合包括该维度上所有向量的元素，所述训练数据集中所有向量的维度相同；

根据预设距离差和量化误差，获取用于划分区间的目标函数；

30 根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间。每一维度对应的区间均是不重叠的。

在实际应用中，根据所述目标函数，将每一维度的元素集合进行区间划分，可为：枚举所述元素集合中区间划分的边界值，将每组所述区间划分的边界值作为所述目标函数的参数值，确定并记录所述目标函数对应的取值；

选择使目标函数取得极值的边界值划分该维度，获得该维度对应的区间。

在更具体的例子中，组合单元43可具体用于，将所有二值码按照每一二值码对应的维度序号顺序拼接，获得二值向量。

5 本实施例的高维向量的哈希装置能够对降低现有技术中近邻查询过程中的存储复杂度，以及提升检索效率。

图5示出了本发明一实施例提供的高维向量的近邻查询装置的结构示意图，如图5所示，本实施例的高维向量的近邻查询装置包括：二值向量获取单元51、哈希二值码获取单元52和相似度度量单元53；

10 其中，二值向量获取单元51，用于获取待处理的高维向量的二值向量；
例如，可采用上述任意实施例所述的高维向量的哈希装置来获取二值向量；

哈希二值码获取单元52用于获取所述二值向量的哈希二值码；

相似度度量单元53用于根据所述哈希二值码的编码方式，选择与该编码方式对应的度量距离，采用所述度量距离度量所述二值向量与其他高维向量对应的二值向量的相似度。

15 上述近邻查询装置可降低现有技术中高额存储复杂度，并且大幅度地提升检索的效率。

可以理解的是，本实施例中的哈希装置和近邻查询装置均在计算设备中实现，如终端、移动终端或服务器等实现向量查询或检索，可应用于图像检索，或向量学习、机器学习等方面。

20 下面结合附图，对发明的向量量化方法及装置进行详细说明。

图6示出了本发明一实施例提供的向量量化方法的流程示意图，如图6所示，本实施例的向量量化方法如下所述。

601、获取待量化的向量中每个维度的元素。

25 举例来说，待量化的 n 维向量为 $X=[x_1, x_2, \dots, x_n]$ ，获取待量化的向量中每个维度的元素 x_i ，其中， $1 \leq i \leq n$ 。

602、根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得所述待量化的向量的量化信息。

其中，所述每个维度对应的区间是根据预设的训练数据集训练获取的。

30 对于每个维度的元素集合中的元素 x_i ，将元素集合中的所有元素量化到该维度对应的 u 个独立的区间 $S_{i1}, S_{i2}, \dots, S_{iu}$ 。其中，该 u 个区间之间不重叠（即区间是独立的，对于一个维度的任意两个区间，它们的交集是空集）。即，若

元素 x_i 属于区间 S_{ij} ，则将元素 x_i 量化到区间 S_{ij} 。

其中， u 表示的是每个维度划分的区间数，比如 $u=4$ ，那么就是将该维度划分成4个区间。

举例来说，在步骤602之前，图6所示的方法还可包括下述的步骤601a；

- 5 601a、获取所述待量化向量的维度，根据所述维度获取与所述维度对应的区间信息，所述区间信息包括：预先获取的所述维度中每个维度对应的区间。

10 例如，待量化向量的维度为64维，则可根据64维获取与64维度对应的区间信息；或者，待量化向量的维度为128维，则可根据128维获取与128维度对应的区间信息。

本实施例的向量量化方法，通过获取待量化的向量中每个维度的元素，进而将元素量化到该元素所属维度对应的区间，获得待量化的向量的量化信息，由此，可实现对高维向量进行量化处理，且采用量化后的信息能够提高图像检索中的检索准确率。

- 15 图7示出了本发明一实施例提供的向量量化方法的流程示意图，如图7所示，本实施例的向量量化方法如下所述。

701、获取待量化的向量中每个维度的元素。

举例来说，待量化的 n 维向量为 $X=[x_1, x_2, \dots, x_n]$ ，获取待量化的向量中每个维度的元素 x_i ，其中， $1 \leq i \leq n$ 。

- 20 702、根据预设的训练数据集训练获取每个维度对应的区间，所述每个维度为所述待量化向量中的每个维度。

在本实施例中，每一维度对应的区间均是不重叠的。

举例来说，步骤701可包括下述的图中未示出的子步骤：

- 25 7021、获取训练数据集中所有向量的每一维度的元素集合，所述元素集合包括该维度上所有向量的元素，所述训练数据集中所有向量的维度相同，且与待量化的向量的维度相同。

例如，选 t 个 n 维向量 $X_1, X_2, \dots, X_t$ 作为训练数据集。对于第 i 个维度，统计这 t 个向量在此维度上所有的取值，不妨设为， $A=\{a_1, a_2, \dots, a_t\}$ 。其中，

$$a_1 \leq a_2 \leq \dots \leq a_t.$$

7022、根据预设距离差和量化误差，获取用于划分区间的目标函数。

7023、根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间。

5 具体地，枚举所述元素集合中区间划分的边界值，将每组所述区间划分的边界值作为所述目标函数的参数值，确定并记录所述目标函数对应的取值；

选择使目标函数取得极值（如最小值）的边界值划分该维度，获得该维度对应的区间。

可理解的是，在具体应用中，目标函数还可能非连续而且不可导的函数，在应用中存在无法直接获取该目标函数的极值，也就是说，无法直接判断该目标函数在某个取值下是否是极大或者极小值。为此，可预先枚举足够的边界值，从中选择使目标函数相对最大（或者最小）的边界值来划分区间。

当然，可根据实际需要选择使目标函数取得最大值或最小值的方式，本实施例均为举例说明，不对其进行限定。

在一个具体的例子中，对于一个维度的元素集合A的区间划分，用一个目标函数来判定该维度的元素集合A的区间划分的好坏。

具体中，假定将元素集合A划分成3个区间。

对于区间划分如下说明：

$$20 \quad A_1 = \{a_1, a_2, \dots, a(l_1)\}$$

$$A_2 = \{a(l_{1+1}), a(l_{1+2}), \dots, a(l_2)\}$$

$$A_3 = \{a(l_{2+1}), \dots, a_t\}.$$

本实施例中采用一个目标函数F判定维度中区间划分质量的好坏。

目标函数的信息如下：

25 当前，
距离差定义成：

$$MSDE = \sum_i \sum_j (d(Q(a_i), Q(a_j)) - t * d(V(a_i), V(a_j)))^2 \dots \dots \dots;$$

量化误差定义成： $MSE = \sum_i (a_i - Q(a_i))^2$;

其中, a_i 被量化到的区间的均值为 $Q(a_i)$ 。即, 如果 a_i 对应的区间为 A_j , $Q(a_i)$ 代表 A_j 中所有元素的平均值。

t 是一个放缩因子, 调整 $d(Q(a_i), Q(a_j))$ 和 $d(V(a_i), V(a_j))$ 的差值的绝对值大小。

5 $V(a_i)$ 代表元素 a_i 所属的向量。即, 如果 a_i 属于 X_j ,那么 $V(a_i)=X_j$ 。

最后, 目标函数由MSDE和MSE的一个组合构成, 即:

$$F = \lambda * MSDE + \mu * MSE$$

λ 和 μ 的取值可根据实际需要选择。举例来说, ($\lambda=1, \mu=0$), ($\lambda=0, \mu=1$), ($\lambda=0.5, \mu=0.5$)等。

10 寻找能够使上述定义的目标函数取得极值的区间的边界值。即, 寻找一个使 F 取得极小值的划分方式。

实际中, 如果 $A_1=\{a_1, a_2, \dots, a(s_1)\}$, $A_2=\{a(s_{1+1}), a(s_{1+2}), \dots, a(s_2)\}$, $A_3=\{a(s_{2+1}), \dots, a_t\}$ 对应的 F 最小, 那么选 $a(s_1)$ 和 $a(s_2)$ 对应的划分方式。

选用上述使目标函数取得极值的划分方案来确定维度的区间。

15 使用上述的 $a(s_1)$ 和 $a(s_2)$ 将维度划分成3段。

在本实施例中, 查找所述元素集合中区间划分的边界值, 将所述区间划分的边界值作为所述目标函数的参数值, 确定所述目标函数的取值是否为极值; 如果所述目标函数的取值为极值, 则采用该区间划分的边界值划分该维度的元素集合中的元素, 获得该维度对应的区间, 举例如下:

20 可枚举每个划分边界的取值, 具体中, 假定将 $A=\{a_1, a_2, \dots, a_t\}$ 划分成3个区间 $A_1=\{a_1, a_2, \dots, a_{(l1)}\}$

$$A_2=\{a_{(l1+1)}, a_{(l1+2)}, \dots, a_{(l2)}\}$$

$$A_3=\{a_{(l2+1)}, \dots, a_t\}。$$

那么, 枚举 $l1$ 和 $l2$ 的取值。 $1 \leq l1 \leq t, l1 < l2 \leq t$ 。

25 对于每一组取值, 带入目标函数计算函数值。当 $l1$ 和 $l2$ 确定后, 计算对应 F 的取值。

枚举过程中, 记录下所有 $l1$ 和 $l2$ 组合对应的 F 的取值。

选所有取值中最优的一组来划分维度。 假如某组 $l1=p, l2=q$ 使 F 最小。

那么选用 p 和 q 将维度划分成三部分，

$$A_1 = \{a_1, a_2, \dots, a_p\}$$

$$A_2 = \{a_{(p+1)}, a_{(p+2)}, \dots, a_q\}$$

$$A_3 = \{a_{(q+1)}, a_{(q+2)}, \dots, a_t\}.$$

- 5 703、根据获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得所述待量化的向量的量化信息。

本实施例中的量化可理解为是一种分组，量化后也就是意味着被分到了特定的组，而量化信息就是所属组的编号或者用来表示该组的符号。

例如，假设是量化第 i 维的元素 y_i ，将第 i 维的元素分为三组， $(-\infty, -1]$ ，
10 $(-1, 1]$ ， $(1, \infty)$ ，每个组编号分别是1, 2, 3。对于一个 $y_i=0.5$ ，它被分到 $(-1, 1]$ ，这个过程是量化。它被编码成2，这就是量化信息。

另外，需要说明的是，原始向量之间的相关性体现在目标函数中的

$$MSDE = \sum_i \sum_j (d(Q(a_i), Q(a_j)) - t * d(V(a_i), V(a_j)))^2$$
 这一部分中。
 在寻找目标函数的极值时，不仅仅依赖对应维度的元素集合中元素的分布，
 15 还会利用维度对应的向量之间的距离关系，为此本实施例中可称之为原始向量之间的相关性。

由此，本实施例中的向量量化方法能够结合原始向量之间的相关性，进而对每一维度的每一元素进行量化，进而在向量的维度很高时也不会破坏向量的空间结构，由此，采用量化后的信息在图像检索中检索准确率提高。

20 另外，本发明实施例还提供一种使用上述图6或图7所示的向量量化方法在图像检索中使用的方法如下：

M01、获取一幅图像，并提取一个描述图像视觉特性的高维向量。

M02、统计高维向量每个维度的数值分布，基于分布信息给每个维度量化。

25 M03、将每个维度量化后的信息进行二值编码，从而得到高维向量的量化信息。

M04、选择与二值编码对应的距离度量方法度量量化后的向量之间的相

似性。

上述方法在图像检索过程中可提高图像检索中的检索准确率。

图8示出了本发明一实施例提供的向量量化装置的结构示意图，如图8所示，本实施例的向量量化装置包括：元素获取单元81、量化单元82；

5 元素获取单元81用于获取待量化的向量中每个维度的元素；

量化单元82用于根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得所述待量化的向量的量化信息；

其中，所述每个维度对应的区间是根据预设的训练数据集训练获取的。

在一种具体的例子中，上述向量量化装置还可包括下述的图中未示出的
10 维度确定单元，该维度确定单元用于获取所述待量化向量的维度，根据所述维度，获取与所述维度对应的区间信息，所述区间信息包括：预先获取的所述维度中每个维度对应的区间。

在另一种具体的例子中，向量量化装置还可包括下述的图中未示出的维度划分单元，该维度划分单元可为离线完成过程，用于根据预设的训练数据集训练获取每个维度对应的区间，所述每个维度为所述待量化向量中的每个
15 维度；

每一维度对应的区间均是不重叠的。

例如，维度划分单元具体用于，获取训练数据集中所有向量的每一维度的元素集合，所述元素集合包括该维度上所有向量的元素，所述训练数据集中所有向量的维度相同，且与待量化的向量的维度相同；
20

根据预设距离差和量化误差，获取用于划分区间的目标函数；

根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间。

或者，维度划分单元具体用于，获取训练数据集中所有向量的每一维度的元素集合，所述元素集合包括该维度上所有向量的元素，所述训练数据集中所有向量的维度相同，且与待量化向量的维度相同；
25

根据预设距离差和量化误差，获取用于划分区间的目标函数；

枚举所述元素集合中区间划分的边界值，将每组所述区间划分的边界值

作为所述目标函数的参数值，确定并记录所述目标函数对应的取值；

选择使目标函数取得极值的边界值划分该维度，获得该维度对应的区间。

本实施例的向量量化装置，可实现对高维向量进行量化处理，且采用量化后的信息能够提高图像检索中的检索准确率。

5 本发明的说明书中，说明了大量具体细节。然而，能够理解，本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中，并未详细示出公知的方法、结构和技术，以便不模糊对本说明书的理解。

类似地，应当理解，为了精简本发明公开并帮助理解各个发明方面中的一个或多个，在上面对本发明的示例性实施例的描述中，本发明的各个特征
10 有时被一起分组到单个实施例、图、或者对其的描述中。然而，并不应将该公开的方法解释呈反映如下意图：即所要求保护的本发明要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说，如下面的权利要求书所反映的那样，发明方面在于少于前面公开的单个实施例的所有特征。因此，遵循具体实施方式的权利要求书由此明确地并入该具体实施方式，其中每个
15 权利要求本身都作为本发明的单独实施例。

本领域技术人员可以理解，可以对实施例中的设备中的模块进行自适应性地改变并且把它们设置在于该实施例不同的一个或多个设备中。可以把实施例中的模块或单元或组件组合成一个模块或单元或组件，以及此外可以把它们分成多个子模块或子单元或子组件。除了这样的特征和/或过程或者单元
20 中的至少一些是互相排斥之处，可以采用任何组合对本说明书（包括伴随的权利要求、摘要和附图）中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述，本说明书（包括伴随的权利要求、摘要和附图）中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

25 此外，本领域的技术人员能够理解，尽管在此所述的一些实施例包括其它实施例中所包括的某些特征而不是其它特征，但是不同实施例的特征的组合意味着处于本发明的范围之内并且形成不同的实施例。例如，在下面的权利要求书中，所要求保护的实施例的任意之一都可以以任意的组合方式来使

用。

本发明的各个部件实施例可以以硬件实现，或者以在一个或者多个处理器上运行的软件模块实现，或者以它们的组合实现。应该注意的是上述实施例对本发明进行说明而不是对本发明进行限制，并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中，不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本发明可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中，
5 这些装置中的若干个可以是通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

最后应说明的是：以上各实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述各实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分或者全部技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的范围，其均应涵盖在本发明的权利要求和说明书的范围当中。
15

权利要求书

1、一种高维向量的哈希方法，其特征在于，包括：

采用标量量化方式对压缩后的高维向量进行量化处理，获取每个元素的量化信息，其中，所述每个元素为压缩后的高维向量中每个维度的元素；

5 将所述每个元素的量化信息编码成二值码；

按照预设规则将所有二值码进行组合，获得二值向量。

2、根据权利要求1所述的方法，其特征在于，采用标量量化方式对压缩后的高维向量进行量化处理，获取每个元素的量化信息，包括：

获取压缩后的高维向量中每个维度的元素；

10 根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得每个元素的量化信息；

所述每个维度对应的区间是根据预设的训练数据集训练获取的。

3、根据权利要求2所述的方法，其特征在于，根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间之前，所述方法
15 还包括：

获取所述压缩后高维向量的维度，根据所述维度获取与所述维度对应的区间信息，所述区间信息包括：预先获取的所述维度中每个维度对应的区间。

4、根据权利要求2或3所述的方法，其特征在于，根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间之前，所述方
20 法还包括：

根据预设的训练数据集训练获取每个维度对应的区间，所述每个维度为所述高维向量中的每个维度；

每一维度对应的区间均是不重叠的。

5、根据权利要求4所述的方法，其特征在于，根据预设的训练数据集训练获取每个维度对应的区间，包括：
25

获取训练数据集中所有向量的每一维度的元素集合，所述元素集合包括该维度上所有向量的元素，所述训练数据集中所有向量的维度相同，均等于

所述压缩后的高维向量的维度；

根据预设距离差和量化误差，获取用于划分区间的目标函数；

根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间。

- 5 6、根据权利要求5所述的方法，其特征在于，根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间，包括：

枚举所述元素集合中区间划分的边界值，将每组所述区间划分的边界值作为所述目标函数的参数值，确定并记录所述目标函数对应的取值；

选择使目标函数取得极值的边界值划分该维度，获得该维度对应的区间。

- 10 7、根据权利要求1至6任一所述的方法，其特征在于，按照预设规则将所有二值码进行组合，获得二值向量，包括：

将所有二值码按照每一二值码对应的维度序号顺序拼接，获得二值向量。

8、一种高维向量的近邻查询方法，其特征在于，包括：

采用权利要求1至7任一所述的方法获取待处理的高维向量的二值向量；

- 15 获取所述二值向量的哈希二值码；

根据所述哈希二值码的编码方式，选择与该编码方式对应的度量距离，采用所述度量距离度量所述二值向量与其他高维向量对应的二值向量的相似度。

9、一种高维向量的哈希装置，其特征在于，包括：

- 20 量化单元，用于采用标量量化方式对压缩后的高维向量进行量化处理，获取每个元素的量化信息，其中，所述每个元素为压缩后的高维向量中每个维度的元素；

编码单元，用于将所述每个元素的量化信息编码成二值码；

组合单元，用于按照预设规则将所有二值码进行组合，获得二值向量。

- 25 10、一种高维向量的近邻查询装置，其特征在于，包括：

二值向量获取单元，用于采用上述权利要求9所述的装置获取待处理的高维向量的二值向量；

哈希二值码获取单元，用于获取所述二值向量的哈希二值码；

相似度度量单元，用于根据所述哈希二值码的编码方式，选择与该编码方式对应的度量距离，采用所述度量距离度量所述二值向量与其他高维向量对应的二值向量的相似度。

11、一种向量量化方法，其特征在于，包括：

5 获取待量化的向量中每个维度的元素；

根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得所述待量化的向量的量化信息；

其中，所述每个维度对应的区间是根据预设的训练数据集训练获取的。

12、根据权利要求11所述的方法，其特征在于，所述根据预先获取的每个
10 维度对应的区间，将所述元素量化到该元素所属维度对应的区间之前，所述方法还包括：

获取所述待量化向量的维度，根据所述维度，获取与所述维度对应的区间信息，所述区间信息包括：预先获取的所述维度中每个维度对应的区间。

13、根据权利要求11或12所述的方法，其特征在于，所述根据预先获取
15 的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间之前，所述方法还包括：

根据预设的训练数据集训练获取每个维度对应的区间，所述每个维度为所述待量化向量中的每个维度；

每一维度对应的区间均是不重叠的。

20 14、根据权利要求13所述的方法，其特征在于，根据预设的训练数据集训练获取每个维度对应的区间，包括：

获取训练数据集中所有向量的每一维度的元素集合，所述元素集合包括该维度上所有向量的元素，所述训练数据集中所有向量的维度相同，且与待量化向量的维度相同；

25 根据预设距离差和量化误差，获取用于划分区间的目标函数；

根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间。

15、根据权利要求14所述的方法，其特征在于，根据所述目标函数，将

每一维度的元素集合进行区间划分，获得该维度对应的区间，包括：

枚举所述元素集合中区间划分的边界值，将每组所述区间划分的边界值作为所述目标函数的参数值，确定并记录所述目标函数对应的取值；

选择使目标函数取得极值的边界值划分该维度，获得该维度对应的区间。

5

16、一种向量量化装置，其特征在于，包括：

元素获取单元，用于获取待量化的向量中每个维度的元素；

量化单元，用于根据预先获取的每个维度对应的区间，将所述元素量化到该元素所属维度对应的区间，获得所述待量化的向量的量化信息；

10

其中，所述每个维度对应的区间是根据预设的训练数据集训练获取的。

17、根据权利要求16所述的装置，其特征在于，所述装置还包括：

维度确定单元，用于获取所述待量化向量的维度，根据所述维度，获取与所述维度对应的区间信息，所述区间信息包括：预先获取的所述维度中每个维度对应的区间。

15

18、根据权利要求16或17所述的装置，其特征在于，所述装置还包括：

维度划分单元，用于根据预设的训练数据集训练获取每个维度对应的区间，所述每个维度为所述待量化向量中的每个维度；

每一维度对应的区间均是不重叠的。

19、根据权利要求18所述的装置，其特征在于，维度划分单元，具体用于

20

获取训练数据集中所有向量的每一维度的元素集合，所述元素集合包括该维度上所有向量的元素，所述训练数据集中所有向量的维度相同，且与待量化向量的维度相同；

根据预设距离差和量化误差，获取用于划分区间的目标函数；

25

根据所述目标函数，将每一维度的元素集合进行区间划分，获得该维度对应的区间。

20、根据权利要求18所述的装置，其特征在于，维度划分单元，具体用于

获取训练数据集中所有向量的每一维度的元素集合，所述元素集合包括该维度上所有向量的元素，所述训练数据集中所有向量的维度相同，且与待量化向量的维度相同；

根据预设距离差和量化误差，获取用于划分区间的目标函数；

- 5 枚举所述元素集合中区间划分的边界值，将每组所述区间划分的边界值作为所述目标函数的参数值，确定并记录所述目标函数对应的取值；

选择使目标函数取得极值的边界值划分该维度，获得该维度对应的区间。

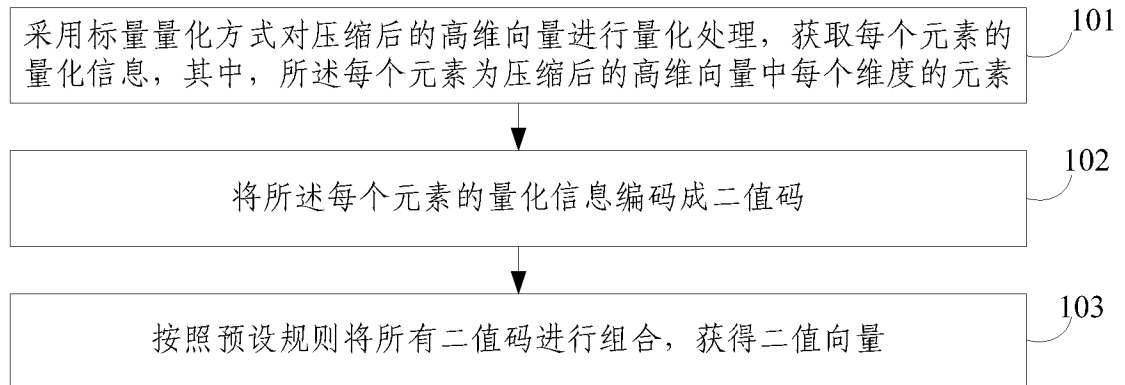


图 1

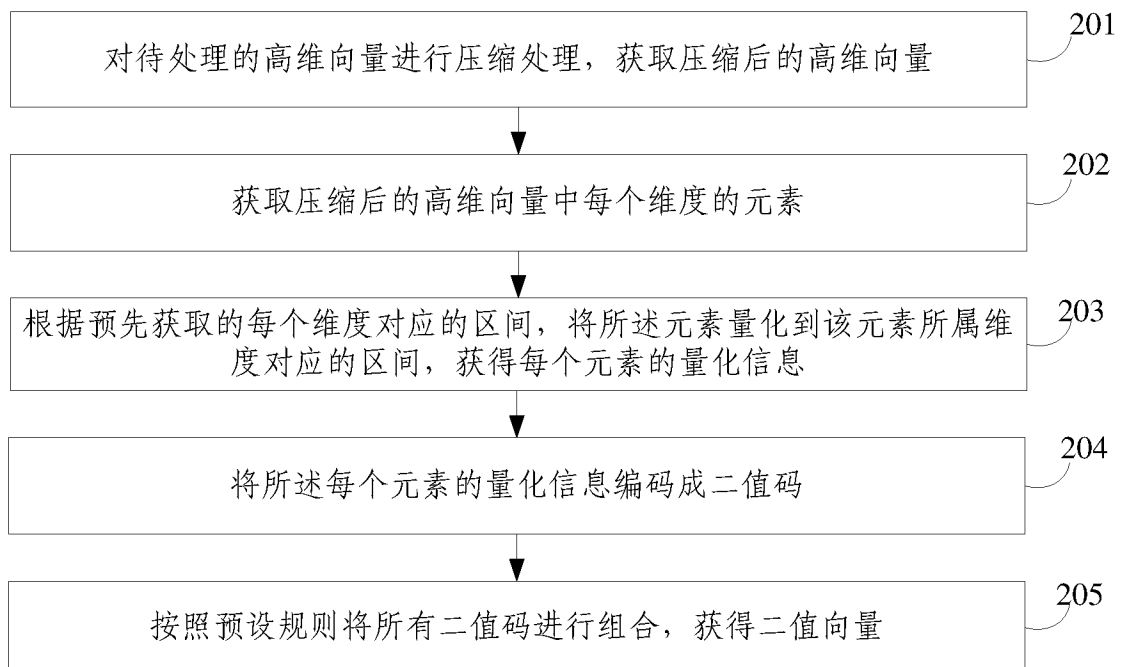


图 2

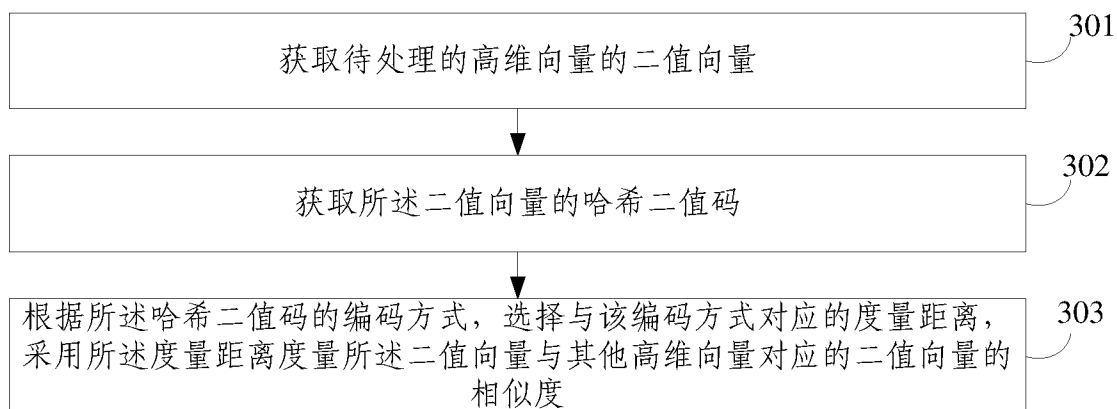


图 3

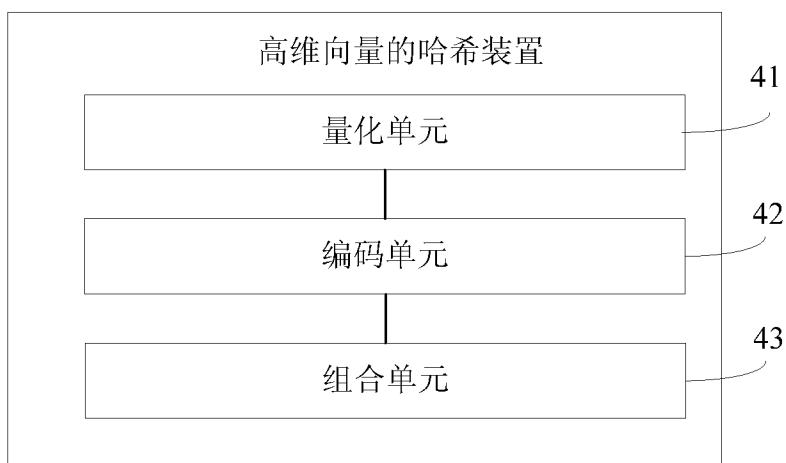


图 4

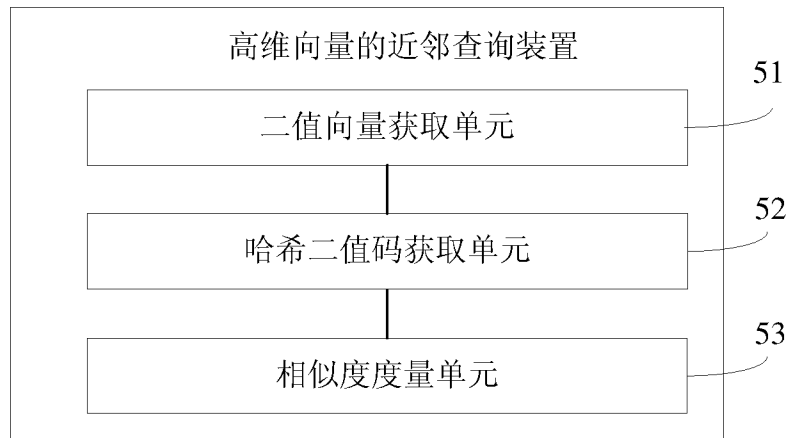


图 5

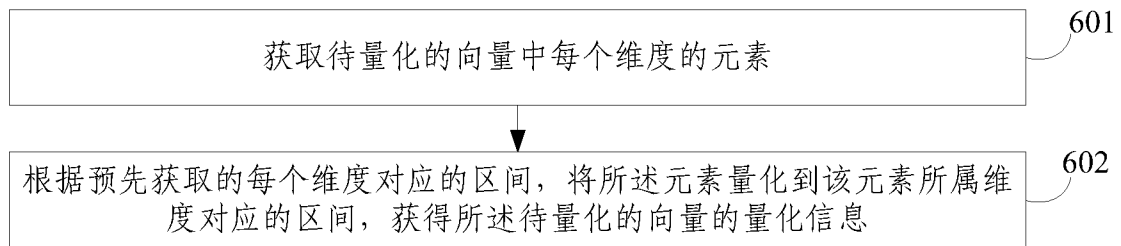


图 6

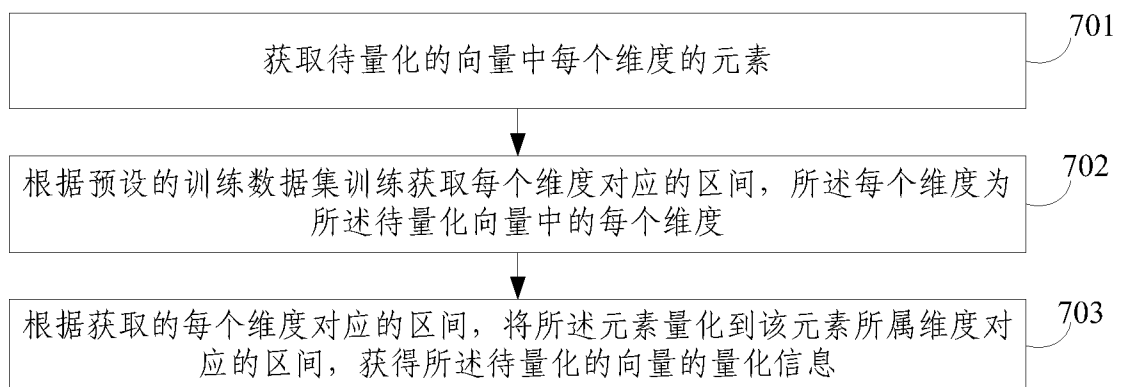


图 7



图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2015/089607

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (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 H04L

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)

WPI, EPODOC, CNPAT, CNKI, IEEE: high dimension, binary encoding, high, dimension, hash, vector, quantization, binary

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101430709 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.), 13 May 2009 (13.05.2009), description, pages 4-28, and figures 1-5	1-20
A	CN 1858773 A (INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES), 08 November 2006 (08.11.2006), the whole document	1-20
A	CN 102819582 A (WASU MEDIA & NETWORK CO., LTD. et al.), 12 December 2012 (12.12.2012), the whole document	1-20

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 02 December 2015 (02.12.2015)	Date of mailing of the international search report 18 December 2015 (18.12.2015)
Name and mailing address of the ISA/CN: 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 JU, Bo Telephone No.: (86-10) 62413661

International application No.
PCT/CN2015/089607

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2015/089607

A. 主题的分类

G06F 17/30(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 H04L

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))

WPI, EPODOC, CNPAT, CNKI, IEEE: 高维, 向量, 哈希, 量化, 维度, 二进制, quantization, binary

二值码, high, dimension, hash, vector,

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
X	CN 101430709 A (腾讯科技深圳有限公司) 2009年 5月 13日 (2009 - 05 - 13) 说明书第4-28页、附图1-5	1-20
A	CN 1858773 A (中国科学院计算技术研究所) 2006年 11月 8日 (2006 - 11 - 08) 全文	1-20
A	CN 102819582 A (华数传媒网络有限公司等) 2012年 12月 12日 (2012 - 12 - 12) 全文	1-20

☐ 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

2015年 12月 2日

国际检索报告邮寄日期

2015年 12月 18日

ISA/CN的名称和邮寄地址

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

授权官员

鞠博

传真号 (86-10)62019451

电话号码 (86-10)62413661

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2015/089607

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	101430709	A	2009年 5月 13日	无	
CN	1858773	A	2006年 11月 8日	无	
CN	102819582	A	2012年 12月 12日	无	

表 PCT/ISA/210 (同族专利附件) (2009年7月)