**一、2018-2019**

**1.论述什么是多媒体？多媒体分析与理解有哪些应用领域和挑战？（10 分）**

答：

多媒体：多媒体是使用不同内容形式的组合的内容，如文本、音频、图像、动画、视频和交互式内容。

应用领域：应用广泛，数据量大，就数字图像视频而言，就广泛应用在安防、教育、通讯、娱乐等产业中，多媒体还应用在图片检索、内容推荐、移动通迅、社交媒体平台、视频网站等。

面对挑战：

（1）如何表示多媒体数据（海量、高维、非结构化）

（2）如何理解多媒体数据（语义鸿沟）

（3）如何挖掘多媒体数据之间的相互关联（关联复杂，关联的协同性与互补性）

（4）如何满足用户多样化的信息需求（海量数据、用户偏好与个性化）

**2.针对文本、音频与图像数据的特征表示方法，请分别列举 2 种典型特征，并分析其优缺点。（15 分）**

答：

文本

1）词频（TF）表示法

优点：用词在文档中的出现频率可表示出一篇文档中的侧重点，便于统计与分析。

缺点：文中多次出现的介词、系动词等也会被统计出较高的权重。

2）潜在语义分析法（LSA）

优点：通过降维，有效解决了一词多义与一义多词的问题。

缺点：仍未解决文档中间的有序性被破坏的问题。

音频

1）过零率

优点：能侧面反映信号在短时帧内的平均频率。

缺点：只关注了短时窗内的幅值信息，频域信息缺失。

2）梅尔频率倒谱系数（MFLL）

优点：对特征进行了去相关和压缩。

缺点：对所有频率段信息进行同等处理，不能突出重要信息。

图像

1）LBP

优点：一定程度上消除了光照变化的问题，具有旋转不变性，计算速度快。

缺点：光照不均时对应的LBP算子会发生变化，同时LBP也丢失了方向信息。

2）SIFT

优点：具有很好的尺度不变性和鲁棒性。

缺点：实时性不高，具有时特征点较少，对边缘光滑的目标无法准确提取特征点。

**3.卷积神经网络中一个典型层通常包含三种基本操作，请回答每种操作的基本含义或类型，所具备的基本特性或者优缺点。（15 分）**

答：

典型层三种基本操作：卷积 — 非线性变换 — 池化

卷积操作：

1）基本含义或类型：卷积操作又称滤波，卷积核函数又称滤波器。在输入图像上的二维卷积，卷积输出称为feature map。

2）基本特征或优缺点：通常会在同一个卷积层使用多个不同的卷积核以学习图像的不同特征。当卷积核输入包含多个通道时，可以把卷积核看成是3D的。

非线性变换：

1）基本含义或类型：先通过Φ(x)将x空间的点转换成z空间的点，而在z空间上得到一个线性的假设，再恢复到原来的x空间中得到一个二次的假设。

2）基本特征或优缺点：优点是有很好的数学性质，缺点是很容易饱和、输出不是0均值的会对梯度产生影响。

池化操作：

1）基本含义或类型：池化函数使用某一位置的相邻位置的总体统计特征来代替网络在该位置的输出。常用的池化函数有最大池化、平均池化。

2）基本特征或优缺点：当输出作出少量平移时，池化能够帮助输入的表示近似不变,即平移不变。减小参数规模，提高统计效率。在全连接层之前采用全局池化，可以保持全连接层节点数量不变，不受输入图像大小的影响。

**4.请说明图像语义理解的基本研究内容，及其各项内容的含义。请针对其中任一项内容，从中选取 1 个典型算法或模型对其具体实现过程进行详细说明。（15 分）**

答：

1）图像语义理解①研究图像中有什么目标，目标之间的相互关系，图像是什么场景的一门学科。②利用人工智能技术让机器能像人一样自动“看懂”外部环境，并可帮助其更好地“说、动、想”。③本质上是学习图像底层特征与高层语义的映射关系。主要研究图像分类、图像标注、图像目标检测、图像语义分割、图像语义描述、视觉问答等内容。

2）

图像分类：给每幅图像一个类别标签（根据各自在图像信息中所反映的不同特征，把不同类别的目标区分开来的图像处理方法。它利用计算机对图像进行定量分析，把图像或图像中的每个像元或区域划归为若干个类别中的某一种，以代替人的视觉判读。）

图像标注：给图像多个语义标签

图像目标检测：对图像中的目标标注一个类别同时用一个紧致框定位目标（输出一系列带有类别标签的目标框）

图像语义分割：给图像多个语义标签且像素级定位

图像语义描述：用自然语言来描述图像内容，即看图说话。

视觉问答：给定图片和图片相关的问题，回答出正确的答案。

3）基于深度学习的目标检测中的R-CNN模型：

1. 首先输入测试图像；

2 利用selective search算法在图像中提取2000个左右的目标候选区域；

3. 将每个候选区域缩放成227x227的大小输入到CNN，采用fc7层特征输出；

4. 基于CNN的候选区域特征表示，输入到SVM进行分类。

**5.详细说明基于 SVD 和 RBM 推荐方法的基本原理，并比较它们的优劣。（15 分）**

答：

**SVD 推荐方法的基本原理**：对于所有用户对所有商品打分，可以建一个矩阵来表示，而这个矩阵是极其稀疏的。而对于任意矩阵都有它的满秩分解。

RU×I=PU×K·QK×I（U为用户数，I为商品数）

然后用R中已知数据去训练P和Q，使得P和Q相乘的结果最好的拟合已知的评分。

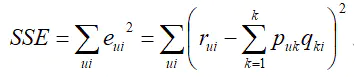
预测用户U对商品的评分：



假设已知的评分r，则预测误差为

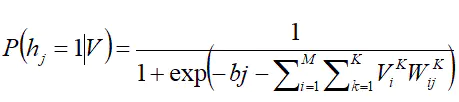


继而可以计算出总的误差平方和：

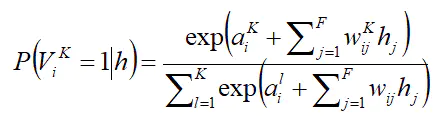


接下来训练的是SSE最小的模型即可。

**RBM 推荐方法的基本原理**：将某一用户对某一商品的评分视作一个softmax神经元，softmax神经元是个长度为k的向量，只有一个分量为1，其余分量为0。而未评分的 部分就可以用全0的softmax神经元表示。这样某个用户的评分可以用矩阵V来表示，在给定可见单元的状态下其激活概率为：



同理，在给定隐单元状态下 ，可见单元的激活率为：



算每一个评分K所对应的激活概率，取所有概率的期望作为预测值。

**比较**：SVD计算过程虽更简单，但由于训练目标单一，容易造成过拟合，而RBM则可以防止梯度爆炸和梯度消失，但求期望的过程会比较复杂，学习效率过慢。

**6.简单描述迭代量化哈希方法(****Iterative Quantization，ITQ)的基本思想,并比较 ITQ 方法与局部敏感哈希(LSH)方法的优劣。（15 分）**

答：

迭代量化哈希方法的基本思想：先对数据集进行PCA降维，然后寻找量化误差最小的旋转矩阵即可得到对应该最优旋转矩阵下的特征向量的二进制编码。（查找 PCA 嵌入式数据；对于c\_bit代码，采用顶部c PCA方向和零阈值）

比较ITQ 方法与局部敏感哈希(LSH)方法的优劣：

1.ITQ

优点：相比局部敏感哈希(LSH)方法多了一步操作，即数据降维后使用矩阵旋转优化，可以降低量化误差

劣势：由于PCA不同维度的方差不平衡，旋转PCA投影数据以尽量减少量化误差时需不断控制旋转角度，即找出最优的旋转矩阵和与之对应的编码，相对麻烦

2.LSH

优点：通过hash function映射变换操作，将原始数据集合分成了多个子集合，而每个子集合中的数据间是相邻的且该子集合中的元素个数较小，因此将一个在超大集合内查找相邻元素的问题转化为了在一个很小的集合内查找相邻元素的问题，计算量下降了很多，改善近似检索的计算表现。

劣势：局部敏感哈希(LSH)方法并不能保证一定能够查找到与query data point最相邻的数据

**7.运动目标检测的困难有哪些？并简述目前常用方法的优缺点。（15 分）**

答：

困难：光照的变化、动态背景、伪装、相机抖动、相机自动对焦、间断的物体运动、阴影

目前常用方法：帧间差方法、背景差分法、统计平均、高斯模型、非参数核密度概率估计、基于码本的运动目标检测、背景建模（ViBe、SuBSENSE）

优缺点：

1.帧间差方法

优点：算法简单、容易实现、检测速度较快。而且一般相邻两帧的时间间隔比较短，因此对

场景的光线变化不是特比敏感。

缺点：对噪声非常敏感且检测出的物体位置不精确，其次帧差法的检测结果与目标运动速度

和相邻两帧间隔大小有关。目标运动过快会被分成两个目标；目运动过慢会被视作背景。

2.背景差分法

优点：算法比较简单；一定程度上克服了环境光线的影响；

缺点：不能用于运动的摄像头；对背景图像实时更新困难；

3.统计平均法

优点：选择适当的参数能够很好地修正背景图像，从而得到一个较为真实的背景估计图像。

缺点：对于频繁出现的运动目标或者在场景内停留的时间较长，该模型不能够很好的提取

运动目标。在复杂场景中会将一些伪目标（如摇摆的树枝等）检测为运动目标。这是由于

在场景中树枝的摇摆会引起像素值的变化。

4.混合高斯模型

优点：能适应背景随时间的缓慢变化，能描述背景中的如显示器屏幕闪烁，树枝晃动等一些周期性扰动等。

缺点：对于缓慢运动的目标不能精确的检测提取，不能够很好的适应阴影、噪声造成的误检或漏检，且不能适应场景的突然变化。

5.非参数核密度概率估计

优点：可以渐进收敛于任意形状的概率密度，而且对动态的场景还具有一定的适应性。

缺点：运算量非常大，很难实现对视频图像的实时检测。对内存要求比较高

6.基于码本的运动目标检测

优点：鲁棒性强，计算效率高，速度快，运算量少，精确度高。

缺点:当训练帧存在较大运动前景时，码本模型的建立会非常不准确，更新参数的调整非常复杂，不能广泛应用于实际情况；由于对于视频每个像素都要建立一个甚至多个码本模型，训练模型比较耗时，若需要背景重构，将大大降低运算速度。

7. ViBe

优点：思想简单，易于实现。（用有限样本近似无限时间窗口）计算量小。运算效率高。（样本少；优化了相似度匹配算法）样本衰减策略。（随机更新策略使得样本生命周期指数衰减，不同于其他方法的先进先出）

缺点：有鬼影、静止目标、阴影前景和运动目标不完整等问题。

8.SuBSENSE

优点：使用了反馈机制，更好适应不同场景，对噪声更为鲁棒。

缺点：（PPT没有）

**二、2019-2020**

**1、论述什么是多媒体?多媒体分析与理解有哪些应用领域和挑战? (10分)**

PPT2前面

**2、简述特征表示学习的必要性，及其主要解决方法与基本原理。(15 分)**

PPT2后面

**3、简述反向误差传播算法的基本原理，并说明该算法在神经网络训练过程中可能遇到的问题及其应对措施。(15 分)**

PPT3前面

**4、简述图像语义理解的基本含义及其面临的主要挑战，选取1个用于图像语义理解任务的典型算法或模型对其具体实现过程进行详细说明。(15分)**

PPT4后面PPT5前面

**5、在NetFlix比赛中，单模型性能最好前两种算法是什么?分别阐述其基本原理及它们之间的优劣。(15 分)**

PPT11后面，上面第5题

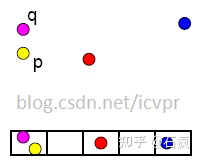
**6、简单描述迭代量化哈希方法(Iterative Quantization, ITQ) 的基本思想，并比较ITQ方法与****局部敏感哈希(LSH) 方法的优劣。(15 分)**

PPT10前面，上面第6题

局部敏感哈希：

一般的hash，原内容发生微小变化后，hash值的变化是无法预估的（字符串改一个字母后，整个字符串md5变得完全不一样；图片改一个像素后hash值也变得完全不一样）局部敏感hash的改进在于，原内容发生微小变化后，其hash值也只发生微小变化。从而满足原内容相近hash值也相近的良好性质。这种性质的好处在于，可以在hash空间进行近邻检索。

我们知道，通过建立Hash Table的方式我们能够得到O(1)的查找时间性能，其中关键在于选取一个hash function，将原始数据映射到相对应的桶内（bucket, hash bin），例如对数据求模：h = x mod w，w通常为一个素数。在对数据集进行hash 的过程中，会发生不同的数据被映射到了同一个桶中（即发生了冲突collision），这一般通过再次哈希将数据映射到其他空桶内来解决。这是普通Hash方法或者叫传统Hash方法，其与LSH有些不同之处。



局部敏感哈希示意图（from: Piotr Indyk）

LSH的基本思想是：将原始数据空间中的两个相邻数据点通过相同的映射或投影变换（projection）后，这两个数据点在新的数据空间中仍然相邻的概率很大，而不相邻的数据点被映射到同一个桶的概率很小。也就是说，如果我们对原始数据进行一些hash映射后，我们希望原先相邻的两个数据能够被hash到相同的桶内，具有相同的桶号。对原始数据集合中所有的数据都进行hash映射后，我们就得到了一个hash table，这些原始数据集被分散到了hash table的桶内，每个桶会落入一些原始数据，属于同一个桶内的数据就有很大可能是相邻的，当然也存在不相邻的数据被hash到了同一个桶内。因此，如果我们能够找到这样一些hash functions，使得经过它们的哈希映射变换后，原始空间中相邻的数据落入相同的桶内的话，那么我们在该数据集合中进行近邻查找就变得容易了，我们只需要将查询数据进行哈希映射得到其桶号，然后取出该桶号对应桶内的所有数据，再进行线性匹配即可查找到与查询数据相邻的数据。换句话说，我们通过hash function映射变换操作，将原始数据集合分成了多个子集合，而每个子集合中的数据间是相邻的且该子集合中的元素个数较小，因此将一个在超大集合内查找相邻元素的问题转化为了在一个很小的集合内查找相邻元素的问题，显然计算量下降了很多。

那具有怎样特点的hash functions才能够使得原本相邻的两个数据点经过hash变换后会落入相同的桶内？这些hash function需要满足以下两个条件：

1）如果d(x,y) ≤ d1， 则h(x) = h(y)的概率至少为p1；

2）如果d(x,y) ≥ d2， 则h(x) = h(y)的概率至多为p2；

其中d(x,y)表示x和y之间的距离，d1 < d2， h(x)和h(y)分别表示对x和y进行hash变换。

满足以上两个条件的hash functions称为(d1,d2,p1,p2)-sensitive。而通过一个或多个(d1,d2,p1,p2)-sensitive的hash function对原始数据集合进行hashing生成一个或多个hash table的过程称为Locality-sensitive Hashing。

**使用LSH进行对海量数据建立索引（Hash table）并通过索引来进行近似最近邻查找的过程如下：**

1. 离线建立索引

（1）选取满足(d1,d2,p1,p2)-sensitive的LSH hash functions；

（2）根据对查找结果的准确率（即相邻的数据被查找到的概率）确定hash table的个数L，每个table内的hash functions的个数K，以及跟LSH hash function自身有关的参数；

（3）将所有数据经过LSH hash function哈希到相应的桶内，构成了一个或多个hash table；

2. 在线查找

（1）将查询数据经过LSH hash function哈希得到相应的桶号；

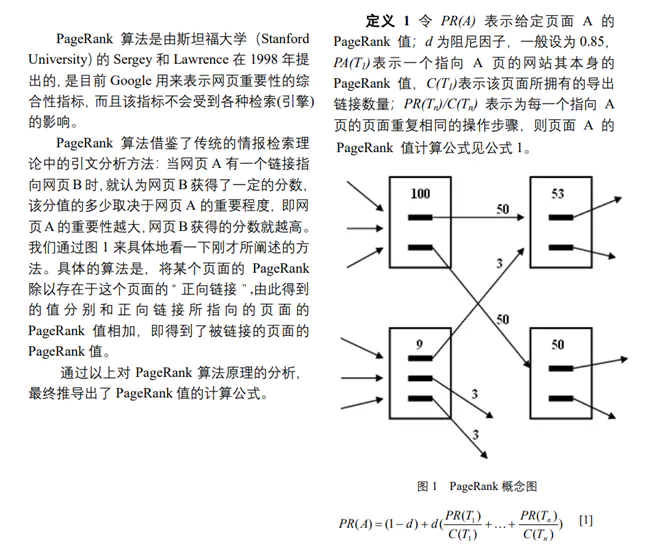
（2）将桶号中对应的数据取出；（为了保证查找速度，通常只需要取出前2L个数据即可）；

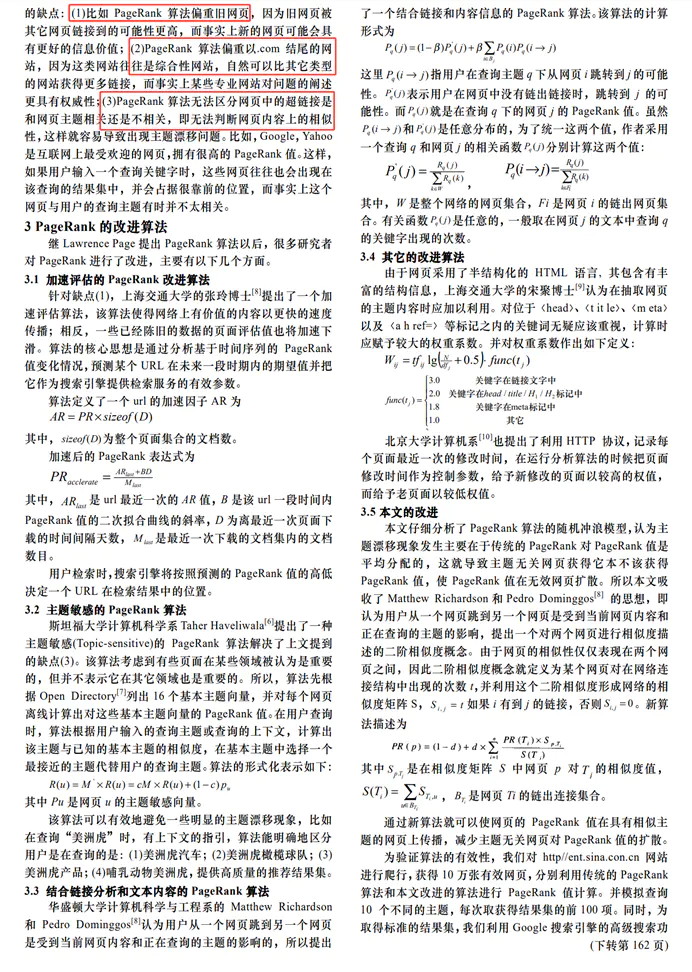
（3）计算查询数据与这2L个数据之间的相似度或距离，返回最近邻的数据；

LSH在线查找时间由两个部分组成： （1）通过LSH hash functions计算hash值（桶号）的时间；（2）将查询数据与桶内的数据进行比较计算的时间。因此，LSH的查找时间至少是一个sublinear时间。为什么是“至少”？因为我们可以通过对桶内的属于建立索引来加快匹配速度，这时第（2）部分的耗时就从O(N)变成了O(logN)或O(1)（取决于采用的索引方法）。

LSH为我们提供了一种在海量的高维数据集中查找与查询数据点（query data point）近似最相邻的某个或某些数据点。需要注意的是，LSH并不能保证一定能够查找到与query data point最相邻的数据，而是减少需要匹配的数据点个数的同时保证查找到最近邻的数据点的概率很大。

**7、阐述PageRank的基本原理，并试着阐述还可能有哪些改进的方式。(15分)**





**三、2020-2021**

**1、论述什么是多媒体?多媒体分析与理解有哪些应用领域和挑战? (15分)**

看往年卷

**2、多媒体内容描述有哪几类代表性的特征表示学习方法，对比说明其异同点，并每类列举2-3种典型算法。(20分)**

PPT2 后面

**3、简述语言模型的基本概念与应用领域，并给出3种典型的语言模型的基本解决思路。(15分)**

PPT4

**4、在NetFlix比赛中，单模型性能最好前两种算法是什么?分别阐述其基本原理及它们之间的优劣。(15 分)**

看往年卷 推荐系统那一章ppt

**5、简单描述****迭代量化哈希方法(Iterative Quantization, ITQ) 的基本思想，并比较ITQ方法与局部敏感哈希(LSH) 方法的优劣。(20 分)**

**Iterative Quantization: A Procrustean Approach to Learning Binary Codes**

**CVPR 2011**

**Yunchao Gong and Svetlana Lazebnik**

目的：

学习保持相似度的二值编码，用于大规模图像采集的高效检索。

贡献：

提出了一种有效的最小化方案来寻找一个旋转的零中心数据，以最小化量化误差。这种方法被称为迭代量化(ITQ)。它既可以用于无监督数据嵌入(如PCA)，也可以用于有监督数据嵌入(如典型相关分析(CCA))。

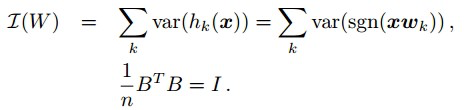
学习二进制码的有效方案：

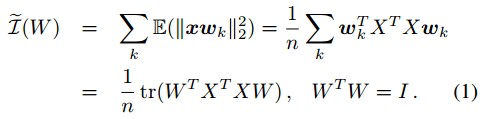
1. 代码应该很短，这样我们就可以在内存中存储大量的图像。
2. 这些代码应该将与具有低汉明距离的二进制字符串相似(无论是感知上还是语义上)的图像映射出来。
3. 学习二进制码参数和编码新的测试图像的算法应该是非常有效的。

方法：

1. 降维

产生一个有效的代码，其中每个比特的方差是最大的。这个目标函数与主成分分析(PCA)完全相同。



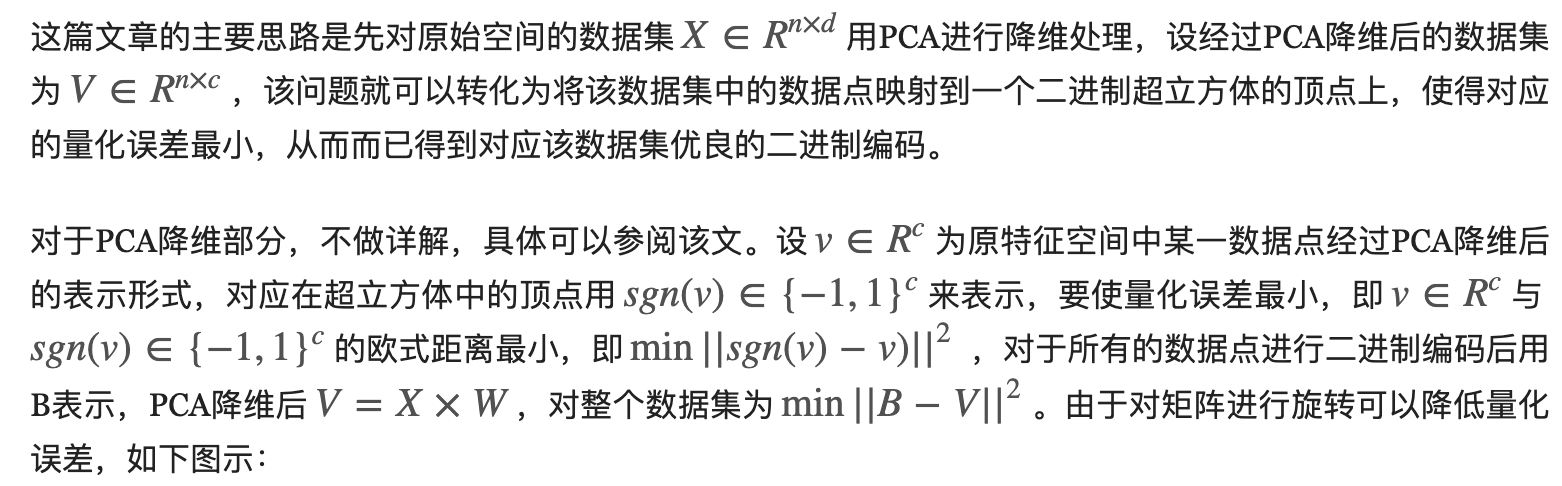


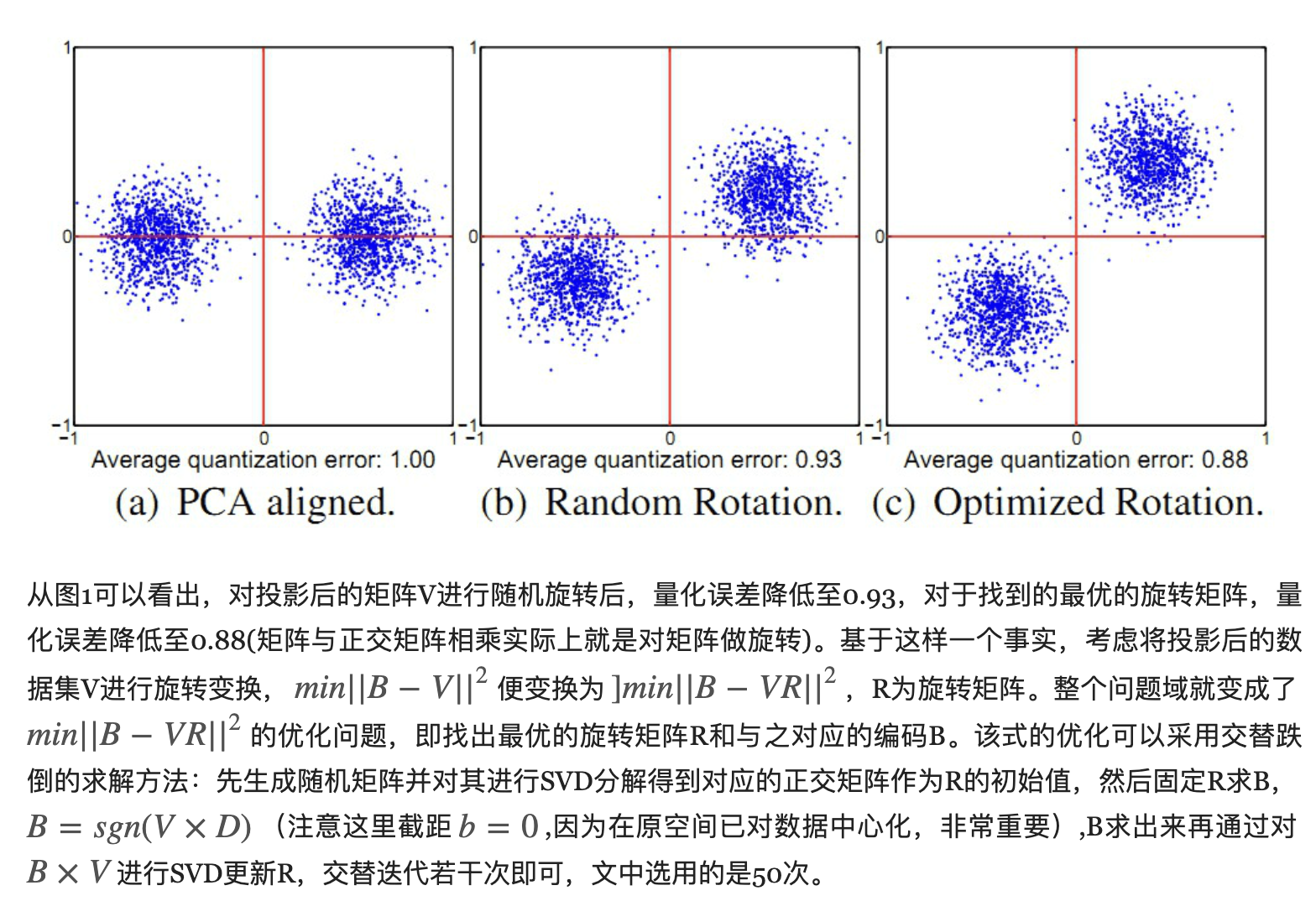
1. 二进制量化

将量化损失最小化，如下所示。



根据量化损失，迭代“修复R更新B”和“修复B更新R”，直到达到局部最小值。





**6、阐述PageRank的基本原理，并试着阐述还可能有哪些改进的方式。(15分)**

看往年卷

**四、PPT目录：**

**第01讲-多媒体内容概述**

**第02讲-多媒体特征表示与提取**

1.多媒体技术概述

2.传统多媒体特征

3.多媒体特征表示学习

**第03讲-基于深度神经网络的特征学习**

1.深度学习概述

2.深度学习基础知识（BatchPerceptron，BackPropagation）

3.典型深度学习模型介绍（自编码，CNN，RNN，Transformer）

**第04讲-自然语言理解**

1.自然语言概述

2.自然语言理解技术

3.几个典型语言模型（N-gram,NNLM,Word2Vec,ELMO,GPT,Bert）

**第05讲-图像视频语义理解**

1.图像语义理解（上一章最后图像分类。图像标注，目标检测，图像语义分割，图像语义描述，视觉问答）

2.视频语义理解：行为分析

**第06讲-cross media analysis**

1.cross-modality, cross-domain/collection, cross-space

(cross-media①Retrieval②Ranking③Hashing④topic modeling⑤other application)

**第07讲-社交媒体分析与理解**

1.社交媒体综述

2.多模态网络表示学习

3.模型安全与数据隐私

**第08讲-基于内容的图像视频检索**

1.困难与挑战

2.各种图像检索与技术分析

**第09讲-视频检索**

困难，内容，镜头检测原理方法，镜头拼接

**第10讲-Rank and Hashing**

1.索引与排序（PageRank，LSH vs ITQ）

2.近似近邻搜索

3.总结展望

**第11讲-多媒体内容推荐-2020.pdf**

1.基于协同过滤的推荐方法CF

2.基于内容的推荐方法

3.推荐系统的评价（SVD/MF vs RBM）