



## 【火炉炼AI】机器学习021-使用K-means进行图片的矢量量化操作



科技老丁哥 关注 IP属地: 北京

2018.09.03 15:01:45 字数 1,815 阅读 925

## 【火炉炼AI】机器学习021-使用K-means进行图片的矢量量化操作

(本文所使用的Python库和版本号: Python 3.5, Numpy 1.14, scikit-learn 0.19, matplotlib 2.2)

前一篇文章我们讲解了K-means算法的定义方法，并用K-means对数据集进行了简单的聚类分析。此处我们讲解使用k-means对图片进行矢量量化操作。

### 1. 矢量量化简介

矢量量化 (Vector Quantization, VQ) 是一种非常重要的信号压缩方法，在图片处理，语音信号处理等领域占据十分重要的地位。

矢量量化是一种基于块编码规则的有损数据压缩方法，在图片压缩格式JPEG和视频压缩格式MPEG-4中都有矢量量化这一步，其基本思想是：将若干个标量数据组构成一个矢量，然后在矢量空间给以整体量化，从而达到压缩数据的同时但不损失多少信息。

矢量量化实际上是一种逼近，其核心思想和“四舍五入”基本一样，就是用一个数来代替其他一个数或者一组数据，比如有很多数据 (6.235, 6.241, 6.238, 6.238954, 6.24205...)，这些数据如果用四舍五入的方式，都可以得到一个数据6.24，即用一个数据 (6.24) 来就可以代表很多个数据。

了解了这个基本思想，我们可以看下面的一维矢量量化的例子：



一维矢量量化举例

在这个数轴上，有很多数据，我们可以用-3来代表所有小于-2的数据，用-1代表-2到0之间的数据，用1代表0到2之间的数据，用3代表大于2的数据，故而整个数轴上的无限多个数据，都可以用这四个数据 (-3, -1, 1, 3) 来表示，我们可以对这四个数进行编码，只需要两个bit就可以，如 (-3=00, -1=01, 1=10, 3=11)，所以这就是1-dimensional, 2-bit VQ，其量化率  $\text{rate} = 2\text{bits/dimension}$ 。

下面看看稍微复杂一点的二维矢量量化的例子：



### 热门故事

我是月照国的女祭司，我的夫君却以为我是个奴婢

我跟女友回家见父母，她小妈居然是我暗恋对象

纪实：女子赴意大利会情郎，海关检查受托带的行李箱，她当场崩溃

演金丝雀太入戏，他还真以为我爱上他了

为了活命，我对病娇反派弟弟表白，他竟当真要做我夫君

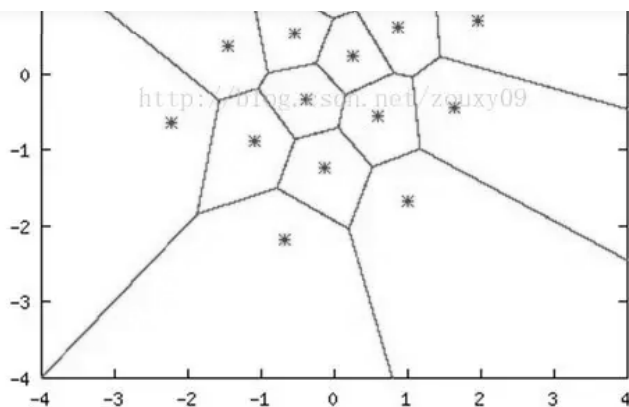
“有个坐过牢的富豪老公是种什么体验？”“要不然你来试试？”

前世渣男把我迷晕还叫我别怕，转世彻底黑化的我复仇反杀

妹妹过失杀人，警察来时，我捡起了那把滴血的刀

我被校霸堵在巷口，却发现他是我谈了三个月的网恋对象

我和网恋对象奔现，却发现他是我们学校赫赫有名的海王



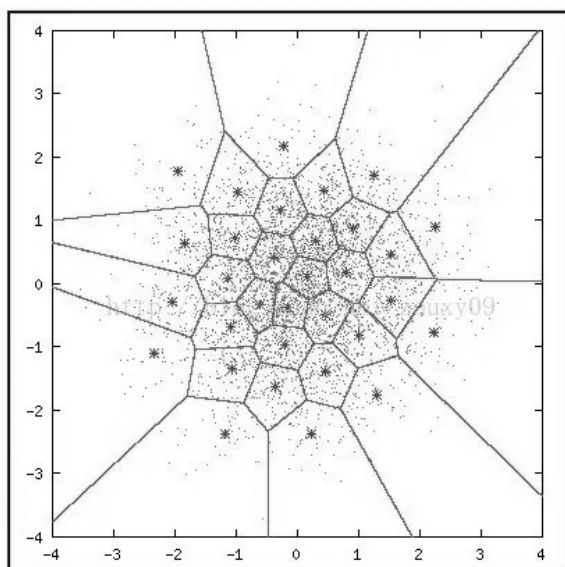
二维矢量量化的例子

由于是二维，故而平面上的任意一个点都可以表示为  $(x,y)$  这种坐标形式，图中，我们用蓝色实线将整个二维平面划分为16个区域，故而任意一个数据点都会落到这16个区域的某一个。我们可以用平面上的某些点来**代表**这个平面区域，故而得到16个红点，这16个红点的坐标就代表了某一个区域内的所有二维点。

更进一步，我们就用4bit二进制码来编码表示这16个数，故而这个问题是2-dimensional, 4-bit VQ，其量化率也是 $\text{rate}=2\text{bits/dimension}$ 。

此处图中显示的红星，也就是16个代表，被称为编码矢量（code vectors），而蓝色边界定的区域叫做编码区域（encoding regions），所有这些编码矢量的集合被称为码书（code book），所有编码区域的集合称为空间的划分（partition of the space）。

对于图像而言，可以认为图像中的每个像素点就是一个数据，用k-means对这些数据进行聚类分析，比如将整幅图像聚为K类，那么会得到K个不同的质心（关于质心的理解和直观感受，可以参考我的上一篇文章【火炉炼AI】机器学习020-使用K-means算法对数据进行聚类分析），或者说通俗一点，可以得到K个不同的数据代表，这些数据代表就可以代表整幅图像中的所有点的像素值，故而我们只需要知道这K个数据代表就可以了（想想人大代表就明白这个道理了），从而可以极大的减少图片的存储空间（比如一张bmp的图像可能有2-3M，而压缩成jpg后只有几百K的大小，当然压缩成jpg的过程还有其他压缩方式，不仅仅是矢量量化，但大体意思相同），当然，这个代表的过程会造成一定的图像像素失真，失真的程度就是K的个数了。用图片可以表示为：



矢量量化压缩图像的简单说明

### 热门故事

我是月照国的女祭司，我的夫君却以为我是个奴婢

我跟女友回家见父母，她小妈居然是我暗恋对象

纪实：女子赴意大利会情郎，海关检查受托带的行李箱，她当场崩溃

演金丝雀太入戏，他还真以为我爱上他了

为了活命，我对病娇反派弟弟表白，他竟当真要做我夫君

“有个坐过牢的富豪老公是种什么体验？”“要不然你来试试？”

前世渣男把我迷晕还叫我别怕，转世彻底黑化的我复仇反杀

妹妹过失杀人，警察来时，我捡起了那把滴血的刀

我被校霸堵在巷口，却发现他是我谈了三个月的网恋对象

我和网恋对象奔现，却发现他是我们学校赫赫有名的海王



## 2. 使用K-means对图像进行矢量量化操作

根据上面第一部分对矢量量化的介绍，我们可以对某一张图片进行矢量量化压缩，可以从图片中提取K个像素代表，然后用这些代表来表示一张图片。具体的代码为：

```
1 from sklearn.cluster import KMeans
2 # 构建一个函数来完成图像的矢量量化操作
3 def image_VQ(image,K_nums): # 貌似很花时间。。
4     # 构建一个KMeans对象
5     kmeans=KMeans(n_clusters=K_nums,n_init=4)
6     # 用这个KMeans对象来训练数据集，此处的数据集就是图像
7     img_data=image.reshape((-1,1))
8     kmeans.fit(img_data)
9     centroids=kmeans.cluster_centers_.squeeze() # 每一个类别的质心
10    labels=kmeans.labels_ # 每一个类别的标记
11    return np.choose(labels,centroids).reshape(image.shape)
```

上面我们先建立一个函数来完成图像的矢量量化压缩操作，这个操作首先建立一个Kmeans对象，然后用这个KMeans对象来训练图像数据，然后提起分类之后的每个类别的质心和标记，并使用这些质心来直接替换原始图像像素，即可得到压缩之后的图像。

为了查看原始图像和压缩后图像，我们将这两幅图都绘制到一行，绘制的函数为：

```
1 # 将原图和压缩图都绘制出来，方便对比查看效果
2 def plot_imgs(raw_img,VQ_img,compress_rate):
3     assert raw_img.ndim==2 and VQ_img.ndim==2, "only plot gray scale images"
4     plt.figure(12,figsize=(25,50))
5     plt.subplot(121)
6     plt.imshow(raw_img,cmap='gray')
7     plt.title('raw_img')
8
9     plt.subplot(122)
10    plt.imshow(VQ_img,cmap='gray')
11    plt.title('VQ_img compress_rate={:.2f}%'.format(compress_rate))
12    plt.show()
```

为了使用方便，我们可以直接将压缩图像函数和显示图像函数封装到一个更高级的函数中，方便我们直接调用和运行，如下所示：

```
1 import cv2
2 def compress_plot_img(img_path,num_bits):
3     assert 1<=num_bits<=8, 'num_bits must be between 1 and 8'
4     K_nums=np.power(2,num_bits)
5
6     # 计算压缩率
7     compression_rate=round(100*(8-num_bits)/8,2)
8     # print('compression rate is {:.2f}%'.format(compression_rate))
9
10    image=cv2.imread(img_path,0) # 读取为灰度图
11    VQ_img=image_VQ(image,K_nums)
12    plot_imgs(image,VQ_img,compression_rate)
```

准备好了各种操作函数之后，我们就可以直接调用compress\_plot\_img()函数来压缩和显示图像，下面是采用三种不同的比特位来压缩得到的图像，可以对比看看效果。

### 热门故事

我是月照国的女祭司，我的夫君却以为我是个女婢

我跟女友回家见父母，她小妈居然是我暗恋对象

纪实：女子赴意大利会情郎，海关检查受托带的行李箱，她当场崩溃

演金丝雀太入戏，他还真以为我爱上他了

为了活命，我对病娇反派弟弟表白，他竟当真要做我夫君

“有个坐过牢的富豪老公是种什么体验？”“要不然你来试试？”

前世渣男把我迷晕还叫我别怕，转世彻底黑化的我复仇反杀

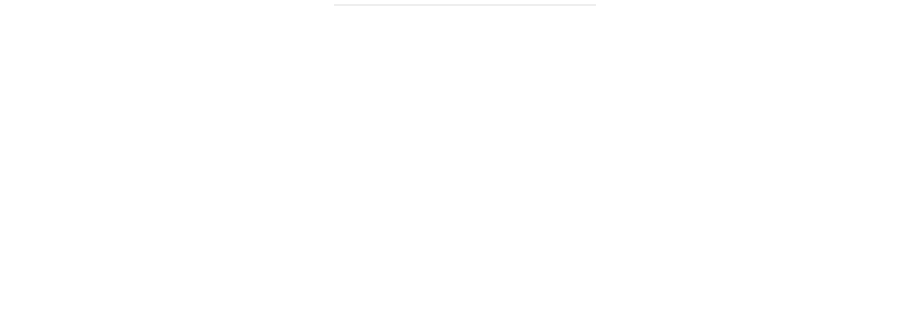
妹妹过失杀人，警察来时，我捡起了那把滴血的刀

我被校霸堵在巷口，却发现他是我谈了三个月的网恋对象

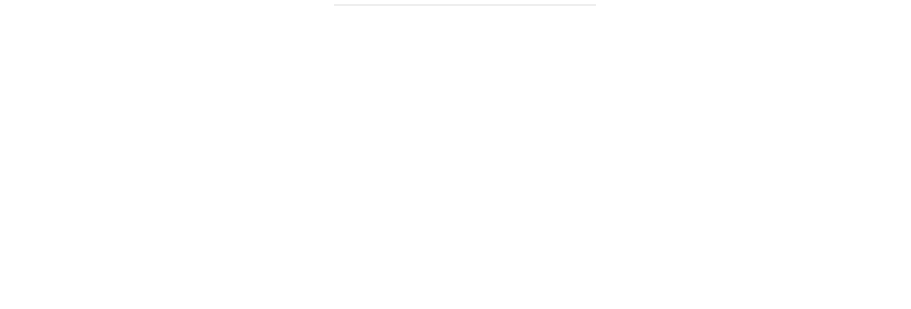
我和网恋对象奔现，却发现他是我们学校赫赫有名的海王



图像压缩率50%的效果对比图



图像压缩率75%的效果对比图



图像压缩率87.5%的效果对比图

#####小\*\*\*\*\*结#####

1, 对图像进行矢量量化压缩, 其本质就是将图像数据划分为K个不同类比, 这种思想和K-means的思想一致, 故而对图像进行矢量量化是 K-means算法的一个重要应用。

2, 通过K-means算法得到图像的K个类别的质心后, 就可以用着 K 个不同质心来代替图像像素, 进而得到压缩之后的, 有少许失真的图像。

3, 从上述三幅图的比较可以看出, 图像压缩率越大, 图像失真的越厉害, 最后的比特位为1时的图像可以说就是二值化图, 其像素值非0即1, 非1即0。

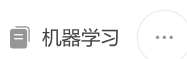
#####

注: 本部分代码已经全部上传到(我的github)上, 欢迎下载。

参考资料:

1, Python机器学习经典实例, Prateek Joshi著, 陶俊杰, 陈小莉译

© 著作权归作者所有. 转载或内容合作请联系作者



更多精彩内容, 就在简书APP



## 热门故事

我是月照国的女祭司, 我的夫君却以为我是个女婢

我跟女友回家见父母, 她小妈居然是我暗恋对象

纪实: 女子赴意大利会情郎, 海关检查受托带的行李箱, 她当场崩溃

演金丝雀太入戏, 他还真以为我爱上他了

为了活命, 我对病娇反派弟弟表白, 他竟当真要做我夫君

“有个坐过牢的富豪老公是种什么体验?” “要不然你来试试?”

前世渣男把我迷晕还叫我别怕, 转世彻底黑化的我复仇反杀

妹妹过失杀人, 警察来时, 我捡起了那把滴血的刀

我被校霸堵在巷口, 却发现他是我谈了三个月的网恋对象

我和网恋对象奔现, 却发现他是我们学校赫赫有名的海王



"小礼物走一走，来简书关注我"

赞赏支持

还没有人赞赏，支持一下



科技老丁哥

总资产12 共写了12.5W字 获得303个赞 共264个粉丝

关注

## 人面猴

序言：七十年代末，一起剥皮案震惊了整个滨河市，随后出现的几起案子，更是在滨河造成了极大的恐慌，老刑警刘岩，带你破解...

沈念sama 阅读 41144 评论 1 赞 87

## 前方高能预警

1. 周嘉洛拥有一个不好说有没有用的异能。异能管理局的人给他的异能起名为「前方高能预警」。2. 周嘉洛第一次发...

沈念sama 阅读 20025 评论 2 赞 77

## 死咒

序言：滨河连续发生了三起死亡事件，死亡现场离奇诡异，居然都是意外死亡，警方通过查阅死者的电脑和手机，发现死者居然都...

沈念sama 阅读 11667 评论 1 赞 81

## 救了他两次的神仙让他今天三更去死

文/潘晓璐 我一进店门，熙熙楼的掌柜王于贵愁眉苦脸地迎上来，“玉大人，你说我怎么就摊上这事。”“怎么了？”我有些...

开封第一讲书人 阅读 8914 评论 0 赞 34

## 道士缉凶录：失踪的卖姜人

文/不坏的土叔 我叫张陵，是天一观的道长。经常有香客问我，道长，这世上最难降的妖魔是什么？我笑而不...

开封第一讲书人 阅读 7297 评论 0 赞 39

## 港岛之恋（遗憾婚礼）

正文 为了忘掉前任，我火速办了婚礼，结果婚礼上，老公的妹妹穿的比我还像新娘。我一直安慰自己，他们只是感情好，可当我...

茶点故事 阅读 10911 评论 0 赞 82

## 恶毒庶女顶嫁案：这布局不是一般人想出来的

文/花漫 我一把揭开白布。她就那样静静地躺着，像睡着了一般。火红的嫁衣衬着肌肤如雪。梳的纹丝不乱的头发上，一...

开封第一讲书人 阅读 10932 评论 0 赞 30

## 热门故事

我是月照国的女祭司，我的夫君却以为我是个女婢

我跟女友回家见父母，她小妈居然是我暗恋对象

纪实：女子赴意大利会情郎，海关检查受托带的行李箱，她当场崩溃

演金丝雀太入戏，他还真以为我爱上他了

为了活命，我对病娇反派弟弟表白，他竟当真要做我夫君

“有个坐过牢的富豪老公是种什么体验？”“要不然你来试试？”

前世渣男把我迷晕还叫我别怕，转世彻底黑化的我复仇反杀

妹妹过失杀人，警察来时，我捡起了那把滴血的刀

我被校霸堵在巷口，却发现他是我谈了三个月的网恋对象


我和网恋对象奔现，却发现他是我们学校赫赫有名的海王



 沈念sama 阅读 4328 评论 2 赞 107

## 双鸳鸯连环套：你想象不到人心有多黑

文/苍兰香墨 我猛地睁开眼，长吁一口气：“原来是场噩梦啊……”“哼！你这毒妇竟也来了？”一声冷哼从身侧响起，我...

 开封第一讲书人 阅读 4063 评论 0 赞 20

## 父亲被人害死了，幕后凶手却是最亲的人！

想象着我的养父在大火中拼命挣扎，窒息，最后皮肤化为焦炭。我心中就已经是抑制不住地欢快，这就叫做以其人之道，还治其人之身...

 爱写小说的胖达 阅读 1636 评论 1 赞 33

## 万荣杀人案实录

序言：老挝万荣一对情侣失踪，失踪者是张志新（化名）和其女友刘颖，没想到半个月后，有当地人在树林里发现了一具尸体，经...

 沈念sama 阅读 4979 评论 0 赞 65

## 护林员之死

正文 独居荒郊野岭守林人离奇死亡，尸身上长有42处带血的胶包..... 初始之章 张勋 以下内容为张勋视角 年9月15日...

 茶点故事 阅读 4913 评论 1 赞 76

## 白月光启示录

正文 我和宋清朗相恋三年，在试婚纱的时候发现自己被绿了。大学时的朋友给我发了我未婚夫和他白月光在一起吃饭的照片。...

 茶点故事 阅读 5289 评论 0 赞 64


## 惨遭霸总抛弃后，我靠赚来的钱成了富豪榜第一名

白月光回国，霸总把我这个替身辞退。还一脸阴沉的警告我。[不要出现在思思面前，不然我有一百种方法让你生不如死。]我...

 爱写小说的胖达 阅读 1081 评论 0 赞 5

## 活死人

序言：一个原本活蹦乱跳的男人离奇死亡，死状恐怖，灵堂内的尸体忽然破棺而出，到底是诈尸还是另有隐情，我是刑警宁泽，带...

 沈念sama 阅读 3766 评论 2 赞 70

## 日本核电站爆炸内幕

正文 年R本政府宣布，位于F岛的核电站，受9级特大地震影响，放射性物质发生泄漏。R本人自食恶果不足惜，却给世界环境...

 茶点故事 阅读 5379 评论 2 赞 65

## 男人毒药：我在死后第九天来索命

文/蒙蒙 一、第九天 我趴在偏房一处隐蔽的房顶上张望。院中可真热闹，春花似锦、人声如沸。这庄子的主人今日做“春日”...

 开封第一讲书人 阅读 3080 评论 0 赞 1

### 热门故事

我是月照国的女祭司，我的夫君却以为我是个女婢

我跟女友回家见父母，她小妈居然是我暗恋对象

纪实：女子赴意大利会情郎，海关检查受托带的行李箱，她当场崩溃

演金丝雀太入戏，他还真以为我爱上他了

为了活命，我对病娇反派弟弟表白，他竟当真要做我夫君

“有个坐过牢的富豪老公是种什么体验？”“要不然你来试试？”

前世渣男把我迷晕还叫我别怕，转世彻底黑化的我复仇反杀

妹妹过失杀人，警察来时，我捡起了那把滴血的刀

我被校霸堵在巷口，却发现他是我谈了三个月的网恋对象

我和网恋对象奔现，却发现他是我们学校赫赫有名的海王





开封第一讲书人 阅读 3199 评论 0 赞 20

## 情欲美人皮

我被黑心中介骗来泰国打工，没想到刚下飞机就差点儿被人妖公主榨干..... 1. 我叫王不留，地道东北人。一个月前我还...

沈念sama 阅读 5307 评论 1 赞 73

## 代替公主和亲

正文 我出身青楼，却偏偏与公主长得像，于是被迫代替她去往敌国和亲。传闻我的和亲对象是个残疾皇子，可洞房花烛夜当晚...

茶点故事 阅读 6608 评论 1 赞 89

## 被以下专题收入，发现更多相似内容

机器学习和人工... 人工智能/模式... 大数据，机器学... sklearn...

## 推荐阅读

更多精彩内容 >

## 【图像检索】【TPAMI重磅综述】SIFT与CNN的碰撞：万字长文回顾图像检索任务...

由于简书对latex公式支持不太友好，因此为了更好的阅读体验，你可以移步至：【TPAMI重磅综述】SIFT与C...

keloli 阅读 5,792 评论 0 赞 18

## 【火炉炼AI】机器学习020-使用K-means算法对数据进行聚类分析

【火炉炼AI】机器学习020-使用K-means算法对数据进行聚类分析 (本文所使用的Python库和版本号: P...

科技老丁哥 阅读 1,574 评论 0 赞 2

## 小猫和她的白肚皮

其实我没有想到你会看我以前写的那些东西，也不知道会给你留下什么样的印象。那时候的我又天真又任性，仗着有人宠爱，行事...

桃溪 阅读 191 评论 5 赞 0

## 李致勋医生的救赎

最近一直盘算着写一些有关《太阳的后裔》中李致勋医生的想法，可忙于仓旗两边折腾，坐公交坐到头痛不说，还把自己的计划打...

文院学渣加油 阅读 525 评论 0 赞 0

## 新年 祈福2019

五福临门！（慈悲不杀生所有动物！）真正的快乐，是大乐！是没有种族歧视，以及让所有的动物都安生无恐。这是小妹的枣...

真如自在 阅读 134 评论 4 赞 3

## 热门故事

我是月照国的女祭司，我的夫君却以为我是个女婢

我跟女友回家见父母，她小妈居然是我暗恋对象

纪实：女子赴意大利会情郎，海关检查受托带的行李箱，她当场崩溃

演金丝雀太入戏，他还真以为我爱上他了

为了活命，我对病娇反派弟弟表白，他竟当真要做我夫君

“有个坐过牢的富豪老公是种什么体验？”“要不然你来试试？”

前世渣男把我迷晕还叫我别怕，转世彻底黑化的我复仇反杀

妹妹过失杀人，警察来时，我捡起了那把滴血的刀

我被校霸堵在巷口，却发现他是我谈了三个月的网恋对象

我和网恋对象奔现，却发现他是我们学校赫赫有名的海王