

# 《大规模多媒体信息管理与检索》课程实验

## ——Label-Region Alignment 系统研发

丁贵广 (教师) 林梓佳 (助教)

2015 年 3 月 31 日

### 1 实验目标

本课程初步计划只设置一个课程实验（相当于集成了往年的两个课程实验和论文阅读训练），要求研发一个针对图像数据的标签-区域对齐（Label-Region Alignment）系统，以巩固课程上所讲解的内容。

该课程实验需要完成的任务包括：图像区域特征的抽取（区域切分已经给出）、SIFT 特征的分布式 K-means 聚类、Label-Region Alignment 算法的实现、网站开发等，考虑到任务划分较为明确，并且总体实验工作量较大，此次实验建议分组完成，每组建议  $\leq 3$  人。本次实验基本上划分为两个阶段，第一阶段主要涉及特征提取，基本上完成到分布式 K-means 聚类；第二阶段主要涉及实现一个 Label-Region Alignment 算法（不限制只能实现给定的算法，可实现其他，但需经过任课老师或者助教的确认）。

此次实验希望能达到如下几个目标：

- 加深同学们对于多媒体数据处理和应用方面的基本技术的了解，巩固课程知识；
- 提升同学们发现问题，解决问题的能力，进而提升研发与编程能力；
- 鼓励同学们团队合作，多思考，多创新，激发对于多媒体信息管理和检索的兴趣。

### 2 实验工具

基本不限制，除特殊要求之外（下文中将详细描述）。

### 3 实验要求

课程实验需要完成的内容详见[章节5](#)。

课程实验最终提交的材料需压缩成 zip 包，按照“组名-组员 1 名字-组员 2 名字-组员 k 名字.zip”的格式提交，每组由组长提交一份即可（根据大家的材料大小情况，届时将通知是提交到网络学堂还是另设的 FTP 服务器），实验材料包括如下：

- source 文件夹

里面包含所有的源码，请根据具体的任务划分子文件夹，并且提供 readme.txt 文件解释各个文件夹的内容，源码要求有恰当的注释；

- document 文件夹

用于提交实验报告，只接受 pdf 版本，实验报告需至少包括如下章节：实验目的、实验环境、实验中引用的材料（诸如开源库等）、**实验基本算法和流程**、实验结果（需根据不同的任务提供截图或者统计数据说明任务实现的结果，要求具备可验证性）、实验结论与心路历程（总结实验过程的经验教训和思考）、组员分工表（描述各自负责的内容，方便组内分数调整）、参考文献。提交的实验报告请尽量做到简明扼要，点明要点为主，无需过分赘述；请使用 IEEE 的模板，下载连接为[http://www.ieee.org/conferences\\_events/conferences/publishing/templates.html](http://www.ieee.org/conferences_events/conferences/publishing/templates.html)，请根据需要选择 word 版本或者 latex 版本，最终提交需转化成 pdf 格式，实验报告请勿超过 12 页。

- exe 文件夹

如果有生成的可执行文件，诸如 exe 或者 jar 包等，可以整理后放入 exe 文件夹中提交，并提供简单的操作说明文件（pdf 格式）。

除此之外，此次实验将划分为两个检查阶段，具体如章节5所述。

## 4 实验评分原则

各小组之间如果存在抄袭现象，抄袭者和被抄袭者所有组员此次实验均为 0 分，第三方代码引用和修改需要在文档中声明，否则视为部分抄袭，将按照抄袭程度扣分，最严重情况同样是此次实验 0 分。

此次实验满分为 100 分，对于实验中出现的具有一定意义的改进和实现，将适当予以加分，但加分额度不超过 5 分，并且实验总分不超过 100 分。

## 5 实验内容

课程实验划分为两个实验阶段，第一个阶段主要是实现特征提取和基于 Hadoop 和 MapReduce 框架（或者 Spark 框架，可自选）的 K-means 聚类，实现的内容包括：搭建 Hadoop（或 Spark）系统、SIFT 点特征抽取【SIFT 特征是必须的，其他特征可自行添加，适当的特征组合一般能有更好的实验效果，建议做成分布式的，但不强制要求】、分布式 K-means 聚类；而第二阶段则是在第一阶段的基础上进一步实现一个 Label-Region Alignment 算法，其主要工作涉及稀疏重构等（具体根据相应的论文而定），最终要求给出一个网站用于直观展示算法的实现效果，同时要求给出数据集上的实验结果文件，用于助教直接计算量化指标（precision 等），以衡量算法效果。

### 5.1 实验阶段一

1. 考察内容：搭建 Hadoop（或者 Spark）系统、图像特征提取（必须包含 SIFT 点特征，可自行多提取其他全局特征或局部特征）、分布式 K-means 聚类
2. 分数比例：50%
3. 完成时间：5 月 9 日，检查将在 5 月 9 日开展，不超过 2 天（如有特殊情况会另行通知，下同），超过提交期限需要跟助教另外约时间检查，并且最终会适当扣分，扣分原则是  $score \times 0.95^n$ ，其中  $score$  是没有扣分前的原始分数， $n$  是迟交的天数
4. 输入：图像数据、每一张图像对应的区域蒙版（mask）

## 5. 实验内容:

- Hadoop 系统搭建【Spark<sup>1</sup>的类似，不做赘述】

请从 Hadoop 官网<sup>2</sup>下载安装包，并根据官方给出的指导文档<sup>3</sup>完成环境搭建工作，非官方指导请自行 Google 或者百度，完成安装后请务必学习 MapReduce 框架的编程模式并用简单的程序测试，考虑到大家实验条件的限制，可考虑使用单机部署 Hadoop 环境，但至少包括 3 个节点。

- SIFT 特征抽取

这部分工作不要求但鼓励分布化实现。SIFT 特征的抽取工具比较多，包括 VLFeat<sup>4</sup> (C、matlab)、javaSIFT<sup>5</sup> (java)、Lire<sup>6</sup> (java) 和 SIFT 作者提供的工具包<sup>7</sup> (C、matlab)，强烈建议使用 VLFeat 或者 SIFT 作者提供的工具包，质量比较高。另外，关于其他特征的使用，可以使用 Lire，请自行学习使用其中的特征提取部分（注意：部分特征的实现并没有在 Lire 中给出，而是在 caliph-emir 项目<sup>8</sup>中，也是 Lire 作者的另一个基础 project，具体可下载阅读）。一般而言，特征的种类越多，如果融合得当，对于图像的描述效果越好，第二阶段的对齐效果也一般会越好。

请注意，对于全局特征，每一个图像只有一个特征向量，而对于 SIFT 等局部特征，每一个图像可能有若干个特征向量（数量不等），所以存放的时候要注意区分，而后面的 K-means 等操作，主要涉及的是局部特征，也就是将数量不等的局部特征量化成一个全局的特征向量，相关的量化方法有 Bag-of-Visual-Words (BoVW)、VLAD、Fisher Vector 等，这里主要是最简单的 BoVW，其他量化方法，同学们感兴趣可以自行查阅，BoVW 的过程将涉及到 K-means 聚类 and Pooling 等操作，具体如下文所述。

- 分布式 K-means 聚类

首先介绍一下 K-means 在 BoVW 中的作用，简单地说，K-means 实际上是为了生成所有 SIFT 点特征的 vocabulary，然后才能将每一个 SIFT 点特征量化到这个 vocabulary 上面，再通过一些类似于 Pooling 的方法将一个图像中的所有 SIFT 点特征融合成一个整体的特征。基本思路如下：

- (a) 确定聚类中心的个数。一般来说，聚类中心个数越多，表达能力越强，但与此同时，由于用来聚类的点特征个数有限，可能导致每个聚类中心分到的点特征数量太少，这样的话，量化是没有意义的，所以聚类中心个数的确定需要综合考虑点特征的数量以及平均下来每个聚类中心所能分到的点特征个数，建议平均下来每个聚类中心的点特征个数要超过 100 个；
- (b) 对所有的点特征进行分布式的 K-means 聚类。这里可以参考 Parallel K-Means Clustering of Remote Sensing Images Based on MapReduce 这篇文章的思路。具体而言，K-means 是一个迭代过程，所以最直接的分布式思路就是对每一轮中的迭代进行分布化，而后重复这个过程，也就是说，K-means 将会有多个 Map-Reduce 过程，每一个迭代对应一个。每一次迭代过程中，Map 过程可以是将上一轮迭代中的聚类中心【每个聚类中心表达成一个特征向量】分发到各个节

---

<sup>1</sup>Spark: <https://spark.apache.org/>

<sup>2</sup>Hadoop: <http://hadoop.apache.org/>

<sup>3</sup>官方文档: <http://hadoop.apache.org/docs/stable/>

<sup>4</sup>VLFeat: <http://www.vlfeat.org/>

<sup>5</sup>javaSIFT: <http://fly.mpi-cbg.de/saalfeld/Projects/javasift.html>

<sup>6</sup>Lire: <http://www.semanticmetadata.net/lire/>

<sup>7</sup>Lowe: <http://www.cs.ubc.ca/lowe/keypoints/>

<sup>8</sup>caliph-emir: <http://sourceforge.net/projects/caliph-emir/files/>

点上，然后在各个节点上计算其所负责的点特征所属的聚类中心，并且记录必须的内容用于 Reduce 过程，而后在 Reduce 的过程中，更新所有聚类中心（也就是聚类中心所对应的特征向量），这里 Map 过程中需要记录的内容留待大家思考。**请注意，这里分布式的 K-means 必须自己实现，不允许直接调用 Spark 或者 Mahout 等项目中现成的 K-means 并行函数，但允许参考实现。**

- (c) 完成图像 SIFT 点特征的量化。重复步骤 2 直到聚类中心的变化很小（具体阈值请自行设置），分布式 K-means 过程完成，并输出所有聚类中心【亦即其所对应的特征向量】。而后则是利用这些聚类中心对每一张图像的所有 SIFT 点特征进行量化。简单的思路是，对于一张图像  $I$ ，其 SIFT 点特征集合假设为  $\{f_1, f_2, \dots, f_t\}$ ，而其最终量化后的向量假定为  $\mathbf{v}$ ，并且  $\mathbf{v}$  的维度等同于聚类中心的个数，那么有  $\mathbf{v}_i = \sum_j \sigma(f_j \in C_i)$ ，其中  $\sigma(\cdot)$  是示性函数，只有条件成立时返回 1，否则返回 0，而  $C_i$  则是第  $i$  个聚类，具体  $f_j \in C_i$  与否，则取决于  $C_i$  是否是距离  $f_j$  最近的聚类中心。这种是 hard assignment 方式，鼓励大家思考其他的赋值方式（可以了解一下 soft assignment 的相关工作）。

所以，基本上每一个区域或者图像（建议先阅读第二阶段的论文再确定）的 SIFT 特征量化完成之后，课程实验的第一阶段即完成。实验检查的要点是特征抽取软件的使用、特征抽取过程与结果、分布式 K-means 的实现、SIFT 点特征的量化等。

## 5.2 实验阶段二

1. 考察内容：实现一个 Label-Region Alignment 算法、网站开发
2. 分数比例：50%
3. 完成时间：6 月 20 日，检查将在 6 月 20 日开始，不超过 2 天（如有特殊情况会另行通知），超过提交期限需要跟助教另外约时间检查，并且最终会适当扣分，扣分原则是  $score \times 0.95^n$ ，其中  $score$  是没有扣分前的原始分数， $n$  是迟交的天数
4. 输入：已经分割好的、带有 label 的图像数据；输出：标记出每一张图像上各个区域所对应的 label。
5. 实验内容：
  - 实现 Label-Region Alignment 算法

首先介绍一下 Label-Region Alignment 的应用场景。在 Label-Region Alignment 中，大家要处理的是整个数据集，数据集中的每一张图像都包含若干个 label，并且为了简化流程，每一张图像中的各个 region 已经给大家分割好了，会有相应的 mask。而大家所要做的事情就是对整个数据集进行处理，将每一张图像的 label 与 region 对应上。所以，这里没有训练集和测试集的概念，只有整个要处理的数据集。因此，Label-Region Alignment 某种意义上讲是对数据集的预处理。另外，对于输入的 mask 文件，其主要由数字组成，基本上就是对一张图像的 region 进行编号，具体则是尺寸与图像大小相同，然后每一个像素对应一个数字，表示其所属的 region 的编号。【所有输入文件的格式在提供的数据集上都有详细的说明】

目前初步确定给大家实现的论文是发表在 ACM MM 2009 上的 Label to Region by Bi-Layer Sparsity Priors，这篇论文的基本思路是通过稀疏重构和标签传播的方式来实现从 image-level label 到 region-level label 的处理过程，其中涉及到稀疏约束等，

大家可通过梯度下降的方法求解（Matlab 中有 SLEP<sup>9</sup>求解包可以用，但是还是建议大家用纯的 java 进行处理，所以大家可以使用普通的梯度下降方式进行求解，以锻炼大家的数学推导能力和编程能力，一般来说，普通梯度下降的效果跟采用针对稀疏约束的梯度下降方法，效果差别不是很大）。具体论文的实现过程中，所遇到的问题可以找助教答疑。另外，正如上文所述，不限制大家实现上述 ACM MM 2009 的论文，大家可选择其他 Label-Region Alignment 的论文进行实现，但是需经过任课老师或者助教的确认。

针对算法的实现效果，要求大家将输出的内容按照指定的格式输出，以便助教进行批量测试，计算标准的 precision 等指标对大家的实现进行评价。具体文件格式说明如下【具体在数据集中会有示例】：每一个图像对应一个文件，与图像的名字相同，并且后缀是.l2r，如“1000.l2r”表示对图像“1000.jpg”的 Label-Region Alignment 结果。文件的内容为文本格式，尺寸与图像大小一致，每一个像素对应一个数字，表示该像素所属的 label【所有的 label 都会进行编号】。

- 交互网站开发

这部分的任务主要是用于直观展示 Label-Region Alignment 的实验效果，具体实现方式可以参考上述论文中的结果展示部分，也可自行设计，以简洁、直观为主。

## 6 附录

- 助教联系方式：linzijia07@tsinghua.org.cn / 13810486125, 东配楼 11 区 417
- 答疑时间：具体请课代表安排时间，集中答疑，课代表信息将公布在网络学堂上
- 相关资源官网：

- Spark: <https://spark.apache.org/>
- Hadoop: <http://hadoop.apache.org/>
- Lire: <http://www.semanticmetadata.net/lire/>
- SLEP: <http://www.yelab.net/software/SLEP/>
- VLFeat: <http://www.vlfeat.org/>
- javaSIFT: <http://fly.mpi-cbg.de/saalfeld/Projects/javasift.html>
- Lire: <http://www.semanticmetadata.net/lire/>
- SIFT 作者提供的工具包: <http://www.cs.ubc.ca/lowe/keypoints/>

---

<sup>9</sup>SLEP: <http://www.yelab.net/software/SLEP/>